

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**



(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
6 December 2001 (06.12.2001)

PCT

(10) International Publication Number  
WO 01/92304 A2(51) International Patent Classification<sup>7</sup>: C07K 14/00

(21) International Application Number: PCT/US01/17065

(22) International Filing Date: 25 May 2001 (25.05.2001)

(25) Filing Language: English

(26) Publication Language: English

## (30) Priority Data:

60/208,424	26 May 2000 (26.05.2000)	US
60/209,001	1 June 2000 (01.06.2000)	US
60/210,588	8 June 2000 (08.06.2000)	US
60/212,335	16 June 2000 (16.06.2000)	US
60/213,747	22 June 2000 (22.06.2000)	US
60/215,391	29 June 2000 (29.06.2000)	US

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

## (72) Inventors; and

(75) Inventors/Applicants (for US only): THORNTON, Michael [US/US]; 9 Medway Road, Woodside, CA 94062 (US). WALIA, Narinder, K. [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). NGUYEN, Dannel, B. [US/US]; 1403 Ridgewood Drive, San Jose, CA 95124 (US). LAI, Preeti [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). GANDHI, Ameena, R. [US/US]; 837 Roble Avenue, #1, Menlo Park, CA 94025 (US). TRIBOULEY, Catherine, M. [FR/US]; 1121 Tennessee Street, #5, San Francisco, CA 94107 (US). YAO, Monique, G. [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). RAMKUMAR, Jayalaxmi [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). AU-YOUNG, Janice [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). LU, Yan [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). AZIMZAI, Yalda [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). BRUNS, Christopher, M. [US/US]; 575 S. Rengstorff Avenue, #126, Mountain View, CA 94040 (US). GRIFFIN, Jennifer, A. [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). YANG, Junming [CH/US]; 7125 Bark Lane, San Jose, CA 95129 (US). SANJANWALA, Madhu, S. [US/US];

210 Sylvia Court, Los Altos, CA 94024 (US). RAU-MANN, Brigitte, E. [US/US]; 5801 South Dorchester Avenue #3B, Chicago, IL 60637 (US). LEE, Ernestine, A. [US/US]; 624 Kains Street, Albany, CA 94706 (US). HAFALIA, April [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). GREEN, Barrie, D. [US/US]; 1332 10th Avenue, #104, San Francisco, CA 94122 (US). KHAN, Farrah, A. [IN/US]; 333 Escuela Avenue, #221, Mountain View, CA 94040 (US). KEARNEY, Liam [IE/US]; 50 Woodside Avenue, San Francisco, CA 94127 (US). ELLIOT, Vicki, S. [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). SEILHAMER, Jeffrey, J. [US/US]; 12555 La Cresta, Los Alots Hills, CA 94022 (US). POLICKY, Jennifer, L. [US/US]; 1511 Jarvis Court, San Jose, CA 95118 (US). BOROWSKY, Mark, L. [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). BURFORD, Neil [GB/US]; 105 Wildwood Circle, Durham, CT 06355 (US). DING, Li [CH/US]; 3353 Alma Street, #146, Palo Alto, CA 94306 (US). LU, Dyung, Aina, M. [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive, #17, Mountain View, CA 94040 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

## Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: TRANSPORTERS AND ION CHANNELS

(57) Abstract: The invention provides human transporters and ion channels (TRICH) and polynucleotides which identify and encode TRICH. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of TRICH.

WO 01/92304 A2

## TRANSPORTERS AND ION CHANNELS

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of transporters and ion channels and to the use of these sequences in the diagnosis, treatment, and prevention of transport, neurological, muscle, immunological, and cell proliferative disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of transporters and ion channels.

### BACKGROUND OF THE INVENTION

Eukaryotic cells are surrounded and subdivided into functionally distinct organelles by hydrophobic lipid bilayer membranes which are highly impermeable to most polar molecules. Cells and organelles require transport proteins to import and export essential nutrients and metal ions including  $K^+$ ,  $NH_4^+$ ,  $P_i$ ,  $SO_4^{2-}$ , sugars, and vitamins, as well as various metabolic waste products. Transport proteins also play roles in antibiotic resistance, toxin secretion, ion balance, synaptic neurotransmission, kidney function, intestinal absorption, tumor growth, and other diverse cell functions (Griffith, J. and C. Sansom (1998) The Transporter Facts Book, Academic Press, San Diego CA, pp. 3-29). Transport can occur by a passive concentration-dependent mechanism, or can be linked to an energy source such as ATP hydrolysis or an ion gradient. Proteins that function in transport include carrier proteins, which bind to a specific solute and undergo a conformational change that translocates the bound solute across the membrane, and channel proteins, which form hydrophilic pores that allow specific solutes to diffuse through the membrane down an electrochemical solute gradient.

Carrier proteins which transport a single solute from one side of the membrane to the other are called uniporters. In contrast, coupled transporters link the transfer of one solute with simultaneous or sequential transfer of a second solute, either in the same direction (symport) or in the opposite direction (antiport). For example, intestinal and kidney epithelium contains a variety of symporter systems driven by the sodium gradient that exists across the plasma membrane. Sodium moves into the cell down its electrochemical gradient and brings the solute into the cell with it. The sodium gradient that provides the driving force for solute uptake is maintained by the ubiquitous  $Na^+/K^+$  ATPase system. Sodium-coupled transporters include the mammalian glucose transporter (SGLT1), iodide transporter (NIS), and multivitamin transporter (SMVT). All three transporters have twelve putative transmembrane segments, extracellular glycosylation sites, and cytoplasmically-oriented N- and C-termini. NIS plays a crucial role in the evaluation, diagnosis, and treatment of various thyroid pathologies because it is the molecular basis for radioiodide thyroid-imaging techniques



and for specific targeting of radioisotopes to the thyroid gland (Levy, O. et al. (1997) Proc. Natl. Acad. Sci. USA 94:5568-5573). SMVT is expressed in the intestinal mucosa, kidney, and placenta, and is implicated in the transport of the water-soluble vitamins, e.g., biotin and pantothenate (Prasad, P.D. et al. (1998) J. Biol. Chem. 273:7501-7506).

5 One of the largest families of transporters is the major facilitator superfamily (MFS), also called the uniporter-symporter-antiporter family. MFS transporters are single polypeptide carriers that transport small solutes in response to ion gradients. Members of the MFS are found in all classes of living organisms, and include transporters for sugars, oligosaccharides, phosphates, nitrates, nucleosides, monocarboxylates, and drugs. MFS transporters found in eukaryotes all have a structure  
10 comprising 12 transmembrane segments (Pao, S.S. et al. (1998) Microbiol. Molec. Biol. Rev. 62:1-34). The largest family of MFS transporters is the sugar transporter family, which includes the seven glucose transporters (GLUT1-GLUT7) found in humans that are required for the transport of glucose and other hexose sugars. These glucose transport proteins have unique tissue distributions and physiological functions. GLUT1 provides many cell types with their basal glucose requirements and  
15 transports glucose across epithelial and endothelial barrier tissues; GLUT2 facilitates glucose uptake or efflux from the liver; GLUT3 regulates glucose supply to neurons; GLUT4 is responsible for insulin-regulated glucose disposal; and GLUT5 regulates fructose uptake into skeletal muscle. Defects in glucose transporters are involved in a recently identified neurological syndrome causing infantile seizures and developmental delay, as well as glycogen storage disease, Fanconi-Bickel syndrome, and  
20 non-insulin-dependent diabetes mellitus (Mueckler, M. (1994) Eur. J. Biochem. 219:713-725; Longo, N. and L.J. Elsas (1998) Adv. Pediatr. 45:293-313).

Monocarboxylate anion transporters are proton-coupled symporters with a broad substrate specificity that includes L-lactate, pyruvate, and the ketone bodies acetate, acetoacetate, and beta-hydroxybutyrate. At least seven isoforms have been identified to date. The isoforms are  
25 predicted to have twelve transmembrane (TM) helical domains with a large intracellular loop between TM6 and TM7, and play a critical role in maintaining intracellular pH by removing the protons that are produced stoichiometrically with lactate during glycolysis. The best characterized H<sup>+</sup>-monocarboxylate transporter is that of the erythrocyte membrane, which transports L-lactate and a wide range of other aliphatic monocarboxylates. Other cells possess H<sup>+</sup>-linked monocarboxylate  
30 transporters with differing substrate and inhibitor selectivities. In particular, cardiac muscle and tumor cells have transporters that differ in their K<sub>m</sub> values for certain substrates, including stereoselectivity for L- over D-lactate, and in their sensitivity to inhibitors. There are Na<sup>+</sup>-monocarboxylate cotransporters on the luminal surface of intestinal and kidney epithelia, which allow the uptake of lactate, pyruvate, and ketone bodies in these tissues. In addition, there are specific and selective

transporters for organic cations and organic anions in organs including the kidney, intestine and liver. Organic anion transporters are selective for hydrophobic, charged molecules with electron-attracting side groups. Organic cation transporters, such as the ammonium transporter, mediate the secretion of a variety of drugs and endogenous metabolites, and contribute to the maintenance of intercellular pH

5 (Poole, R.C. and A.P. Halestrap (1993) *Am. J. Physiol.* 264:C761-C782; Price, N.T. et al. (1998) *Biochem. J.* 329:321-328; and Martinelle, K. and I. Haggstrom (1993) *J. Biotechnol.* 30:339-350).

Recently, Yamashita et al. (Yamashita, T. et al. (1997) *J. Biol. Chem.* 272:10205-10211) have identified a peptide /histidine transporter (PHT1) in rat, expressed particularly in brain and retina tissue. When expressed in *Xenopus* oocytes, PHT1 induces proton-dependent histidine transport. This

10 transport process was inhibited by dipeptides and tripeptides but not free amino acids such as glutamate, glycine, leucine, methionine, and aspartate. This transporter is believed to be a member of a superfamily of proton-coupled peptide and nitrate transporters.

ATP-binding cassette (ABC) transporters are members of a superfamily of membrane proteins that transport substances ranging from small molecules such as ions, sugars, amino acids,

15 peptides, and phospholipids, to lipopeptides, large proteins, and complex hydrophobic drugs. ABC transporters consist of four modules: two nucleotide-binding domains (NBD), which hydrolyze ATP to supply the energy required for transport, and two membrane-spanning domains (MSD), each containing six putative transmembrane segments. These four modules may be encoded by a single gene, as is the case for the cystic fibrosis transmembrane regulator (CFTR), or by separate genes.

20 When encoded by separate genes, each gene product contains a single NBD and MSD. These "half-molecules" form homo- and heterodimers, such as Tap1 and Tap2, the endoplasmic reticulum-based major histocompatibility (MHC) peptide transport system. Several genetic diseases are attributed to defects in ABC transporters, such as the following diseases and their corresponding proteins: cystic fibrosis (CFTR, an ion channel), adrenoleukodystrophy (adrenoleukodystrophy protein, ALDP),

25 Zellweger syndrome (peroxisomal membrane protein-70, PMP70), and hyperinsulinemic hypoglycemia (sulfonylurea receptor, SUR). Overexpression of the multidrug resistance (MDR) protein, another ABC transporter, in human cancer cells makes the cells resistant to a variety of cytotoxic drugs used in chemotherapy (Taglicht, D. and S. Michaelis (1998) *Meth. Enzymol.* 292:130-162).

A number of metal ions such as iron, zinc, copper, cobalt, manganese, molybdenum, selenium,

30 nickel, and chromium are important as cofactors for a number of enzymes. For example, copper is involved in hemoglobin synthesis, connective tissue metabolism, and bone development, by acting as a cofactor in oxidoreductases such as superoxide dismutase, ferroxidase (ceruloplasmin), and lysyl oxidase. Copper and other metal ions must be provided in the diet, and are absorbed by transporters in the gastrointestinal tract. Plasma proteins transport the metal ions to the liver and other target organs,

where specific transporters move the ions into cells and cellular organelles as needed. Imbalances in metal ion metabolism have been associated with a number of disease states (Danks, D.M. (1986) J. Med. Genet. 23:99-106).

Transport of fatty acids across the plasma membrane can occur by diffusion, a high capacity, low affinity process. However, under normal physiological conditions a significant fraction of fatty acid transport appears to occur via a high affinity, low capacity protein-mediated transport process. Fatty acid transport protein (FATP), an integral membrane protein with four transmembrane segments, is expressed in tissues exhibiting high levels of plasma membrane fatty acid flux, such as muscle, heart, and adipose. Expression of FATP is upregulated in 3T3-L1 cells during adipose conversion, and expression in COS7 fibroblasts elevates uptake of long-chain fatty acids (Hui, T.Y. et al. (1998) J. Biol. Chem. 273:27420-27429).

Mitochondrial carrier proteins are transmembrane-spanning proteins which transport ions and charged metabolites between the cytosol and the mitochondrial matrix. Examples include the ADP, ATP carrier protein; the 2-oxoglutarate/malate carrier; the phosphate carrier protein; the pyruvate carrier; the dicarboxylate carrier which transports malate, succinate, fumarate, and phosphate; the tricarboxylate carrier which transports citrate and malate; and the Grave's disease carrier protein, a protein recognized by IgG in patients with active Grave's disease, an autoimmune disorder resulting in hyperthyroidism. Proteins in this family consist of three tandem repeats of an approximately 100 amino acid domain, each of which contains two transmembrane regions (Stryer, L. (1995) Biochemistry, W.H. Freeman and Company, New York NY, p. 551; PROSITE PDOC00189 Mitochondrial energy transfer proteins signature; Online Mendelian Inheritance in Man (OMIM) \*275000 Graves Disease).

This class of transporters also includes the mitochondrial uncoupling proteins, which create proton leaks across the inner mitochondrial membrane, thus uncoupling oxidative phosphorylation from ATP synthesis. The result is energy dissipation in the form of heat. Mitochondrial uncoupling proteins have been implicated as modulators of thermoregulation and metabolic rate, and have been proposed as potential targets for drugs against metabolic diseases such as obesity (Ricquier, D. et al. (1999) J. Int. Med. 245:637-642).

### **Ion Channels**

The electrical potential of a cell is generated and maintained by controlling the movement of ions across the plasma membrane. The movement of ions requires ion channels, which form ion-selective pores within the membrane. There are two basic types of ion channels, ion transporters and gated ion channels. Ion transporters utilize the energy obtained from ATP hydrolysis to actively transport an ion against the ion's concentration gradient. Gated ion channels allow passive flow of an

ion down the ion's electrochemical gradient under restricted conditions. Together, these types of ion channels generate, maintain, and utilize an electrochemical gradient that is used in 1) electrical impulse conduction down the axon of a nerve cell, 2) transport of molecules into cells against concentration gradients, 3) initiation of muscle contraction, and 4) endocrine cell secretion.

## 5 Ion Transporters

Ion transporters generate and maintain the resting electrical potential of a cell. Utilizing the energy derived from ATP hydrolysis, they transport ions against the ion's concentration gradient.

These transmembrane ATPases are divided into three families. The phosphorylated (P) class ion transporters, including  $\text{Na}^+\text{-K}^+$  ATPase,  $\text{Ca}^{2+}\text{-ATPase}$ , and  $\text{H}^+\text{-ATPase}$ , are activated by a

10 phosphorylation event. P-class ion transporters are responsible for maintaining resting potential distributions such that cytosolic concentrations of  $\text{Na}^+$  and  $\text{Ca}^{2+}$  are low and cytosolic concentration of  $\text{K}^+$  is high. The vacuolar (V) class of ion transporters includes  $\text{H}^+$  pumps on intracellular organelles, such as lysosomes and Golgi. V-class ion transporters are responsible for generating the low pH within the lumen of these organelles that is required for function. The coupling factor (F) class  
15 consists of  $\text{H}^+$  pumps in the mitochondria. F-class ion transporters utilize a proton gradient to generate ATP from ADP and inorganic phosphate ( $\text{P}_i$ ).

The P-ATPases are hexamers of a 100 kD subunit with ten transmembrane domains and several large cytoplasmic regions that may play a role in ion binding (Scarborough, G.A. (1999) Curr. Opin. Cell Biol. 11:517-522). The V-ATPases are composed of two functional domains: the  $\text{V}_1$   
20 domain, a peripheral complex responsible for ATP hydrolysis; and the  $\text{V}_0$  domain, an integral complex responsible for proton translocation across the membrane. The F-ATPases are structurally and evolutionarily related to the V-ATPases. The F-ATPase  $\text{F}_0$  domain contains 12 copies of the c subunit, a highly hydrophobic protein composed of two transmembrane domains and containing a single buried carboxyl group in TM2 that is essential for proton transport. The V-ATPase  $\text{V}_0$  domain  
25 contains three types of homologous c subunits with four or five transmembrane domains and the essential carboxyl group in TM4 or TM3. Both types of complex also contain a single a subunit that may be involved in regulating the pH dependence of activity (Forgac, M. (1999) J. Biol. Chem. 274:12951-12954).

The resting potential of the cell is utilized in many processes involving carrier proteins and  
30 gated ion channels. Carrier proteins utilize the resting potential to transport molecules into and out of the cell. Amino acid and glucose transport into many cells is linked to sodium ion co-transport (symport) so that the movement of  $\text{Na}^+$  down an electrochemical gradient drives transport of the other molecule up a concentration gradient. Similarly, cardiac muscle links transfer of  $\text{Ca}^{2+}$  out of the cell with transport of  $\text{Na}^+$  into the cell (antiport).

### Gated Ion Channels

Gated ion channels control ion flow by regulating the opening and closing of pores. The ability to control ion flux through various gating mechanisms allows ion channels to mediate such diverse signaling and homeostatic functions as neuronal and endocrine signaling, muscle contraction, fertilization, and regulation of ion and pH balance. Gated ion channels are categorized according to the manner of regulating the gating function. Mechanically-gated channels open their pores in response to mechanical stress; voltage-gated channels (e.g., Na<sup>+</sup>, K<sup>+</sup>, Ca<sup>2+</sup>, and Cl<sup>-</sup> channels) open their pores in response to changes in membrane potential; and ligand-gated channels (e.g., acetylcholine-, serotonin-, and glutamate-gated cation channels, and GABA- and glycine-gated chloride channels) open their pores in the presence of a specific ion, nucleotide, or neurotransmitter. The gating properties of a particular ion channel (i.e., its threshold for and duration of opening and closing) are sometimes modulated by association with auxiliary channel proteins and/or post translational modifications, such as phosphorylation.

Mechanically-gated or mechanosensitive ion channels act as transducers for the senses of touch, hearing, and balance, and also play important roles in cell volume regulation, smooth muscle contraction, and cardiac rhythm generation. A stretch-inactivated channel (SIC) was recently cloned from rat kidney. The SIC channel belongs to a group of channels which are activated by pressure or stress on the cell membrane and conduct both Ca<sup>2+</sup> and Na<sup>+</sup> (Suzuki, M. et al. (1999) J. Biol. Chem. 274:6330-6335).

The pore-forming subunits of the voltage-gated cation channels form a superfamily of ion channel proteins. The characteristic domain of these channel proteins comprises six transmembrane domains (S1-S6), a pore-forming region (P) located between S5 and S6, and intracellular amino and carboxy termini. In the Na<sup>+</sup> and Ca<sup>2+</sup> subfamilies, this domain is repeated four times, while in the K<sup>+</sup> channel subfamily, each channel is formed from a tetramer of either identical or dissimilar subunits. The P region contains information specifying the ion selectivity for the channel. In the case of K<sup>+</sup> channels, a GYG tripeptide is involved in this selectivity (Ishii, T.M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11651-11656).

Voltage-gated Na<sup>+</sup> and K<sup>+</sup> channels are necessary for the function of electrically excitable cells, such as nerve and muscle cells. Action potentials, which lead to neurotransmitter release and muscle contraction, arise from large, transient changes in the permeability of the membrane to Na<sup>+</sup> and K<sup>+</sup> ions. Depolarization of the membrane beyond the threshold level opens voltage-gated Na<sup>+</sup> channels. Sodium ions flow into the cell, further depolarizing the membrane and opening more voltage-gated Na<sup>+</sup> channels, which propagates the depolarization down the length of the cell. Depolarization also opens voltage-gated potassium channels. Consequently, potassium ions flow

outward, which leads to repolarization of the membrane. Voltage-gated channels utilize charged residues in the fourth transmembrane segment (S4) to sense voltage change. The open state lasts only about 1 millisecond, at which time the channel spontaneously converts into an inactive state that cannot be opened irrespective of the membrane potential. Inactivation is mediated by the channel's N-terminus, which acts as a plug that closes the pore. The transition from an inactive to a closed state requires a return to resting potential.

Voltage-gated Na<sup>+</sup> channels are heterotrimeric complexes composed of a 260 kDa pore-forming  $\alpha$  subunit that associates with two smaller auxiliary subunits,  $\beta$ 1 and  $\beta$ 2. The  $\beta$ 2 subunit is an integral membrane glycoprotein that contains an extracellular Ig domain, and its association with  $\alpha$  and  $\beta$ 1 subunits correlates with increased functional expression of the channel, a change in its gating properties, as well as an increase in whole cell capacitance due to an increase in membrane surface area (Isom, L.L. et al. (1995) Cell 83:433-442).

Non voltage-gated Na<sup>+</sup> channels include the members of the amiloride-sensitive Na<sup>+</sup> channel/degenerin (NaC/DEG) family. Channel subunits of this family are thought to consist of two transmembrane domains flanking a long extracellular loop, with the amino and carboxyl termini located within the cell. The NaC/DEG family includes the epithelial Na<sup>+</sup> channel (ENaC) involved in Na<sup>+</sup> reabsorption in epithelia including the airway, distal colon, cortical collecting duct of the kidney, and exocrine duct glands. Mutations in ENaC result in pseudohypoaldosteronism type 1 and Liddle's syndrome (pseudohyperaldosteronism). The NaC/DEG family also includes the recently characterized H<sup>+</sup>-gated cation channels or acid-sensing ion channels (ASIC). ASIC subunits are expressed in the brain and form heteromultimeric Na<sup>+</sup>-permeable channels. These channels require acid pH fluctuations for activation. ASIC subunits show homology to the degenerins, a family of mechanically-gated channels originally isolated from *C. elegans*. Mutations in the degenerins cause neurodegeneration. ASIC subunits may also have a role in neuronal function, or in pain perception, since tissue acidosis causes pain (Waldmann, R. and M. Lazdunski (1998) Curr. Opin. Neurobiol. 8:418-424; Eglen, R.M. et al. (1999) Trends Pharmacol. Sci. 20:337-342).

K<sup>+</sup> channels are located in all cell types, and may be regulated by voltage, ATP concentration, or second messengers such as Ca<sup>2+</sup> and cAMP. In non-excitable tissue, K<sup>+</sup> channels are involved in protein synthesis, control of endocrine secretions, and the maintenance of osmotic equilibrium across membranes. In neurons and other excitable cells, in addition to regulating action potentials and repolarizing membranes, K<sup>+</sup> channels are responsible for setting resting membrane potential. The cytosol contains non-diffusible anions and, to balance this net negative charge, the cell contains a Na<sup>+</sup>-K<sup>+</sup> pump and ion channels that provide the redistribution of Na<sup>+</sup>, K<sup>+</sup>, and Cl<sup>-</sup>. The pump actively transports Na<sup>+</sup> out of the cell and K<sup>+</sup> into the cell in a 3:2 ratio. Ion channels in the plasma membrane

allow  $K^+$  and  $Cl^-$  to flow by passive diffusion. Because of the high negative charge within the cytosol,  $Cl^-$  flows out of the cell. The flow of  $K^+$  is balanced by an electromotive force pulling  $K^+$  into the cell, and a  $K^+$  concentration gradient pushing  $K^+$  out of the cell. Thus, the resting membrane potential is primarily regulated by  $K^+$  flow (Salkoff, L. and T. Jegla (1995) *Neuron* 15:489-492).

5 Potassium channel subunits of the Shaker-like superfamily all have the characteristic six transmembrane/1 pore domain structure. Four subunits combine as homo- or heterotetramers to form functional K channels. These pore-forming subunits also associate with various cytoplasmic  $\beta$  subunits that alter channel inactivation kinetics. The Shaker-like channel family includes the voltage-gated  $K^+$  channels as well as the delayed rectifier type channels such as the human ether-a-go-go related gene (HERG) associated with long QT, a cardiac dysrhythmia syndrome (Curran, M.E. (1998) *Curr. Opin. Biotechnol.* 9:565-572; Kaczorowski, G.J. and M.L. Garcia (1999) *Curr. Opin. Chem. Biol.* 3:448-458).

A second superfamily of  $K^+$  channels is composed of the inward rectifying channels (Kir). Kir channels have the property of preferentially conducting  $K^+$  currents in the inward direction. These proteins consist of a single potassium selective pore domain and two transmembrane domains, which correspond to the fifth and sixth transmembrane domains of voltage-gated  $K^+$  channels. Kir subunits also associate as tetramers. The Kir family includes ROMK1, mutations in which lead to Bartter syndrome, a renal tubular disorder. Kir channels are also involved in regulation of cardiac pacemaker activity, seizures and epilepsy, and insulin regulation (Doupnik, C.A. et al. (1995) *Curr. Opin. Neurobiol.* 5:268-277; Curran, supra).

The recently recognized TWIK  $K^+$  channel family includes the mammalian TWIK-1, TREK-1 and TASK proteins. Members of this family possess an overall structure with four transmembrane domains and two P domains. These proteins are probably involved in controlling the resting potential in a large set of cell types (Duprat, F. et al. (1997) *EMBO J* 16:5464-5471).

25 The voltage-gated  $Ca^{2+}$  channels have been classified into several subtypes based upon their electrophysiological and pharmacological characteristics. L-type  $Ca^{2+}$  channels are predominantly expressed in heart and skeletal muscle where they play an essential role in excitation-contraction coupling. T-type channels are important for cardiac pacemaker activity, while N-type and P/Q-type channels are involved in the control of neurotransmitter release in the central and peripheral nervous system. The L-type and N-type voltage-gated  $Ca^{2+}$  channels have been purified and, though their functions differ dramatically, they have similar subunit compositions. The channels are composed of three subunits. The  $\alpha_1$  subunit forms the membrane pore and voltage sensor, while the  $\alpha_2\delta$  and  $\beta$  subunits modulate the voltage-dependence, gating properties, and the current amplitude of the channel. These subunits are encoded by at least six  $\alpha_1$ , one  $\alpha_2\delta$ , and four  $\beta$  genes. A fourth subunit,  $\gamma$ , has

been identified in skeletal muscle (Walker, D. et al. (1998) *J. Biol. Chem.* 273:2361-2367; McCleskey, E.W. (1994) *Curr. Opin. Neurobiol.* 4:304-312).

The transient receptor family (Trp) of calcium ion channels are thought to mediate capacitative calcium entry (CCE). CCE is the  $\text{Ca}^{2+}$  influx into cells to resupply  $\text{Ca}^{2+}$  stores depleted by the action of inositol triphosphate (IP3) and other agents in response to numerous hormones and growth factors. Trp and Trp-like were first cloned from *Drosophila* and have similarity to voltage gated  $\text{Ca}^{2+}$  channels in the S3 through S6 regions. This suggests that Trp and/or related proteins may form mammalian CCC entry channels (Zhu, X. et al. (1996) *Cell* 85:661-671; Boulay, G. et al. (1997) *J. Biol. Chem.* 272:29672-29680). Melastatin is a gene isolated in both the mouse and human, and whose expression in melanoma cells is inversely correlated with melanoma aggressiveness in vivo. The human cDNA transcript corresponds to a 1533-amino acid protein having homology to members of the Trp family. It has been proposed that the combined use of malastatin mRNA expression status and tumor thickness might allow for the determination of subgroups of patients at both low and high risk for developing metastatic disease (Duncan, L.M. et al (2001) *J. Clin. Oncol.* 19:568-576).

Chloride channels are necessary in endocrine secretion and in regulation of cytosolic and organelle pH. In secretory epithelial cells,  $\text{Cl}^-$  enters the cell across a basolateral membrane through an  $\text{Na}^+$ ,  $\text{K}^+/\text{Cl}^-$  cotransporter, accumulating in the cell above its electrochemical equilibrium concentration. Secretion of  $\text{Cl}^-$  from the apical surface, in response to hormonal stimulation, leads to flow of  $\text{Na}^+$  and water into the secretory lumen. The cystic fibrosis transmembrane conductance regulator (CFTR) is a chloride channel encoded by the gene for cystic fibrosis, a common fatal genetic disorder in humans. CFTR is a member of the ABC transporter family, and is composed of two domains each consisting of six transmembrane domains followed by a nucleotide-binding site. Loss of CFTR function decreases transepithelial water secretion and, as a result, the layers of mucus that coat the respiratory tree, pancreatic ducts, and intestine are dehydrated and difficult to clear. The resulting blockage of these sites leads to pancreatic insufficiency, "meconium ileus", and devastating "chronic obstructive pulmonary disease" (Al-Awqati, Q. et al. (1992) *J. Exp. Biol.* 172:245-266).

The voltage-gated chloride channels (CLC) are characterized by 10-12 transmembrane domains, as well as two small globular domains known as CBS domains. The CLC subunits probably function as homotetramers. CLC proteins are involved in regulation of cell volume, membrane potential stabilization, signal transduction, and transepithelial transport. Mutations in CLC-1, expressed predominantly in skeletal muscle, are responsible for autosomal recessive generalized myotonia and autosomal dominant myotonia congenita, while mutations in the kidney channel CLC-5 lead to kidney stones (Jentsch, T.J. (1996) *Curr. Opin. Neurobiol.* 6:303-310).

Ligand-gated channels open their pores when an extracellular or intracellular mediator binds to



the channel. Neurotransmitter-gated channels are channels that open when a neurotransmitter binds to their extracellular domain. These channels exist in the postsynaptic membrane of nerve or muscle cells. There are two types of neurotransmitter-gated channels. Sodium channels open in response to excitatory neurotransmitters, such as acetylcholine, glutamate, and serotonin. This opening causes an influx of  $\text{Na}^+$  and produces the initial localized depolarization that activates the voltage-gated channels and starts the action potential. Chloride channels open in response to inhibitory neurotransmitters, such as  $\gamma$ -aminobutyric acid (GABA) and glycine, leading to hyperpolarization of the membrane and the subsequent generation of an action potential. Neurotransmitter-gated ion channels have four transmembrane domains and probably function as pentamers (Jentsch, supra). Amino acids in the second transmembrane domain appear to be important in determining channel permeation and selectivity (Sather, W.A. et al. (1994) *Curr. Opin. Neurobiol.* 4:313-323).

Ligand-gated channels can be regulated by intracellular second messengers. For example, calcium-activated  $\text{K}^+$  channels are gated by internal calcium ions. In nerve cells, an influx of calcium during depolarization opens  $\text{K}^+$  channels to modulate the magnitude of the action potential (Ishi et al., supra). The large conductance (BK) channel has been purified from brain and its subunit composition determined. The  $\alpha$  subunit of the BK channel has seven rather than six transmembrane domains in contrast to voltage-gated  $\text{K}^+$  channels. The extra transmembrane domain is located at the subunit N-terminus. A 28-amino-acid stretch in the C-terminal region of the subunit (the "calcium bowl" region) contains many negatively charged residues and is thought to be the region responsible for calcium binding. The  $\beta$  subunit consists of two transmembrane domains connected by a glycosylated extracellular loop, with intracellular N- and C-termini (Kaczorowski, supra; Vergara, C. et al. (1998) *Curr. Opin. Neurobiol.* 8:321-329).

Cyclic nucleotide-gated (CNG) channels are gated by cytosolic cyclic nucleotides. The best examples of these are the cAMP-gated  $\text{Na}^+$  channels involved in olfaction and the cGMP-gated cation channels involved in vision. Both systems involve ligand-mediated activation of a G-protein coupled receptor which then alters the level of cyclic nucleotide within the cell. CNG channels also represent a major pathway for  $\text{Ca}^{2+}$  entry into neurons, and play roles in neuronal development and plasticity. CNG channels are tetramers containing at least two types of subunits, an  $\alpha$  subunit which can form functional homomeric channels, and a  $\beta$  subunit, which modulates the channel properties. All CNG subunits have six transmembrane domains and a pore forming region between the fifth and sixth transmembrane domains, similar to voltage-gated  $\text{K}^+$  channels. A large C-terminal domain contains a cyclic nucleotide binding domain, while the N-terminal domain confers variation among channel subtypes (Zufall, F. et al. (1997) *Curr. Opin. Neurobiol.* 7:404-412).

The activity of other types of ion channel proteins may also be modulated by a variety of

intracellular signalling proteins. Many channels have sites for phosphorylation by one or more protein kinases including protein kinase A, protein kinase C, tyrosine kinase, and casein kinase II, all of which regulate ion channel activity in cells. Kir channels are activated by the binding of the G $\beta\gamma$  subunits of heterotrimeric G-proteins (Reimann, F. and F.M. Ashcroft (1999) *Curr. Opin. Cell. Biol.* 11:503-508).

- 5 Other proteins are involved in the localization of ion channels to specific sites in the cell membrane. Such proteins include the PDZ domain proteins known as MAGUKs (membrane-associated guanylate kinases) which regulate the clustering of ion channels at neuronal synapses (Craven, S.E. and D.S. Brecht (1998) *Cell* 93:495-498).

### Disease Correlation

- 10 The etiology of numerous human diseases and disorders can be attributed to defects in the transport of molecules across membranes. Defects in the trafficking of membrane-bound transporters and ion channels are associated with several disorders, e.g., cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, von Gierke disease, and certain forms of diabetes mellitus. Single-gene defect diseases resulting in an inability to transport small molecules across
- 15 membranes include, e.g., cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease (van't Hoff, W.G. (1996) *Exp. Nephrol.* 4:253-262; Talente, G.M. et al. (1994) *Ann. Intern. Med.* 120:218-226; and Chillon, M. et al. (1995) *New Engl. J. Med.* 332:1475-1480).

- Human diseases caused by mutations in ion channel genes include disorders of skeletal muscle, cardiac muscle, and the central nervous system. Mutations in the pore-forming subunits of
- 20 sodium and chloride channels cause myotonia, a muscle disorder in which relaxation after voluntary contraction is delayed. Sodium channel myotonias have been treated with channel blockers. Mutations in muscle sodium and calcium channels cause forms of periodic paralysis, while mutations in the sarcoplasmic calcium release channel, T-tubule calcium channel, and muscle sodium channel cause malignant hyperthermia. Cardiac arrhythmia disorders such as the long QT syndromes and
- 25 idiopathic ventricular fibrillation are caused by mutations in potassium and sodium channels (Cooper, E.C. and L.Y. Jan (1998) *Proc. Natl. Acad. Sci. USA* 96:4759-4766). All four known human idiopathic epilepsy genes code for ion channel proteins (Berkovic, S.F. and I.E. Scheffer (1999) *Curr. Opin. Neurology* 12:177-182). Other neurological disorders such as ataxias, hemiplegic migraine and hereditary deafness can also result from mutations in ion channel genes (Jen, J. (1999) *Curr. Opin.*
- 30 *Neurobiol.* 9:274-280; Cooper, *supra*).

Several genetic diseases are attributed to defects in ABC transporters, such as the following diseases and their corresponding proteins: cystic fibrosis (CFTR, an ion channel), adrenoleukodystrophy (adrenoleukodystrophy protein, ALDP), Zellweger syndrome (peroxisomal membrane protein-70, PMP70), congenital hyperbilirubinemia (MOAT), Stargart's disease, which

causes defective vision in children (RIM/ABCR) and hyperinsulinemic hypoglycemia (sulfonylurea receptor, SUR) (Holland, B. and Blight, M.A. (1999) *J. Mol. Biol.* 293:381-399). Overexpression of the multidrug resistance (MDR) protein in human cancer cells makes the cells resistant to a variety of cytotoxic drugs used in chemotherapy (Taglicht, D. and Michaelis, S. (1998) *Meth. Enzymol.* 292:131-163).

Two monomeric ABC transporters have been identified in the human peroxisome membrane: the adrenoleukodystrophy protein (ALDP) and the 70-kDa peroxisomal membrane protein (PMP70). Mutations in the adrenoleukodystrophy gene cause X-linked adrenoleukodystrophy, an inborn error of peroxisomal  $\beta$ -oxidation of very long chain fatty acids. Mutations in the PMP70 genes have been found in patients with Zellweger syndrome, an inborn error of peroxisome biogenesis. The sulfonylurea receptor, an ABC transporter, regulates the function of pancreatic ATP-sensitive  $K^+$  channels, and sulphonylureas are widely used to treat non-insulin dependent diabetes mellitus (Demolombe, S. and Escande, D. (1996) *Trends Pharmacol. Sci.* 17:273-275). Multidrug-resistance (MDR) results from overproduction of another member of the ABC transporter family,

P-glycoprotein. MDR is primarily caused by increased drug extrusion from the resistant cells by P-glycoprotein. The P-glycoproteins have 2 homologous halves, each with 6 hydrophobic segments adjacent to a consensus sequence for nucleotide binding. The hydrophobic segments may form a membrane channel, whereas the nucleotide binding site may be involved in providing energy for drug transport (Saurin, W. et al. (1994) *Mol. Microbiol.* 12:993-1004; Shani, N., et al. (1996) *J. Biol. Chem.* 271:8725-8730; and Koster, W., and Bohm, B. (1992) *Mol. & Gen. Genet.* 232:399-407).

Ion channels have been the target for many drug therapies. Neurotransmitter-gated channels have been targeted in therapies for treatment of insomnia, anxiety, depression, and schizophrenia. Voltage-gated channels have been targeted in therapies for arrhythmia, ischemic stroke, head trauma, and neurodegenerative disease (Taylor, C.P. and L.S. Narasimhan (1997) *Adv. Pharmacol.* 39:47-98). Various classes of ion channels also play an important role in the perception of pain, and thus are potential targets for new analgesics. These include the vanilloid-gated ion channels, which are activated by the vanilloid capsaicin, as well as by noxious heat. Local anesthetics such as lidocaine and mexiletine which blockade voltage-gated  $Na^+$  channels have been useful in the treatment of neuropathic pain (Eglen, *supra*).

Ion channels in the immune system have recently been suggested as targets for immunomodulation. T-cell activation depends upon calcium signaling, and a diverse set of T-cell specific ion channels has been characterized that affect this signaling process. Channel blocking agents can inhibit secretion of lymphokines, cell proliferation, and killing of target cells. A peptide antagonist of the T-cell potassium channel Kv1.3 was found to suppress delayed-type hypersensitivity

and allogenic responses in pigs, validating the idea of channel blockers as safe and efficacious immunosuppressants (Cahalan, M.D. and K.G. Chandy (1997) Curr. Opin. Biotechnol. 8:749-756).

The discovery of new transporters and ion channels and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of transport, neurological, muscle, immunological, and cell proliferative disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of transporters and ion channels.

### SUMMARY OF THE INVENTION

10 The invention features purified polypeptides, transporters and ion channels, referred to collectively as "TRICH" and individually as "TRICH-1," "TRICH-2," "TRICH-3," "TRICH-4," "TRICH-5," "TRICH-6," "TRICH-7," "TRICH-8," "TRICH-9," "TRICH-10," "TRICH-11," "TRICH-12," "TRICH-13," "TRICH-14," "TRICH-15," "TRICH-16," "TRICH-17," "TRICH-18," "TRICH-19," "TRICH-20," "TRICH-21," "TRICH-22," "TRICH-23," "TRICH-24," "TRICH-25,"  
15 "TRICH-26," and "TRICH-27." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of a polypeptide having an amino acid sequence selected  
20 from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-27.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from  
25 the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of a polypeptide  
30 having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-27. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:28-54.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter

sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a  
5 biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

10 The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a  
15 biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

20 Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of a polypeptide having an  
25 amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27.

The invention further provides an isolated polynucleotide selected from the group consisting of  
30 a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. The method comprises a) combining the polypeptide with at least one test compound

under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:28-54, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, ii) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, ii) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54, iii) a polynucleotide



complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

### BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

### DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a

reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

### DEFINITIONS

"TRICH" refers to the amino acid sequences of substantially purified TRICH obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of TRICH. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of TRICH either by directly interacting with TRICH or by acting on components of the biological pathway in which TRICH participates.

An "allelic variant" is an alternative form of the gene encoding TRICH. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding TRICH include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as TRICH or a polypeptide with at least one functional characteristic of TRICH. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding TRICH, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding TRICH. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent TRICH.

Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of TRICH is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of TRICH. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of TRICH either by directly interacting with TRICH or by acting on components of the biological pathway in which TRICH participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind TRICH polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used

to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as  
5 phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring  
10 nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic"  
15 refers to the capability of the natural, recombinant, or synthetic TRICH, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement,  
20 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding TRICH or fragments of TRICH may be  
25 employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to  
30 repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended

and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
10	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
15	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
20	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
25	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide.

Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

5 A "fragment" is a unique portion of TRICH or the polynucleotide encoding TRICH which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15,  
10 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the  
15 specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:28-54 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:28-54, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:28-54 is useful, for  
20 example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:28-54 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:28-54 and the region of SEQ ID NO:28-54 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-27 is encoded by a fragment of SEQ ID NO:28-54. A  
25 fragment of SEQ ID NO:1-27 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-27. For example, a fragment of SEQ ID NO:1-27 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-27. The precise length of a fragment of SEQ ID NO:1-27 and the region of SEQ ID NO:1-27 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the  
30 intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two

or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Reward for match: 1*

*Penalty for mismatch: -2*

*Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 11*

*Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Open Gap: 11 and Extension Gap: 1 penalties*

*Gap x drop-off: 50*

*Expect: 10*



Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

“Human artificial chromosomes” (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term “humanized antibody” refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

“Hybridization” refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the “washing” step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY;

specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C<sub>0</sub>t or R<sub>0</sub>t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of TRICH which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of TRICH which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of TRICH. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological,

functional, or immunological properties of TRICH.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an TRICH may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of TRICH.

"Probe" refers to nucleic acid sequences encoding TRICH, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for

example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs  
5 can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to  
10 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase  
15 sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may  
20 also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved  
25 oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

30 A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have

been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing TRICH, nucleic acids encoding TRICH, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides

by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater

sequence identity over a certain defined length. A variant may be described as, for example, an “allelic” (as defined above), “splice,” “species,” or “polymorphic” variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass “single nucleotide polymorphisms” (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A “variant” of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

## THE INVENTION

The invention is based on the discovery of new human transporters and ion channels (TRICH), the polynucleotides encoding TRICH, and the use of these compositions for the diagnosis, treatment, or prevention of transport, neurological, muscle, immunological, and cell proliferative disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the

polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are transporters and ion channels. For example, SEQ ID NO:1 is 88% identical to rat ABC transporter (GenBank ID g2982567) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0 (scores are rounded down to zero if they are extremely small, e.g. less than  $10^{-300}$ ), which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains an ABC transporter active site domain and transmembrane domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Results from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:1 is an ABC transporter. In an alternative example, SEQ ID NO:4 is 87% identical to human mitochondrial ornithine transporter (GenBank ID g5565862) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $8.1e-141$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:4 also contains a mitochondrial carrier proteins domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:4 is a mitochondrial carrier protein. In an alternative example, SEQ ID NO:8 is 88% identical to rat peptide/histidine transporter (GenBank ID g2208839)



as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $1.8e-262$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:8 also contains a PTR2 proton-dependent oligopeptide transport (POT) family peptide transporter signature as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLAST-DOMO, BLAST-PRODOM, BLIMPS, and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:8 is a transmembrane PTR2 POT family transporter. In an alternative example, SEQ ID NO:15 is 51% identical from amino acid residues 117 to 742 to rat sodium/glucose cotransporter (GenBank ID g286259) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $8.9e-174$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:15 also contains a sodium:solute symporter family domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:15 is a sodium/glucose cotransporter. In an alternative example, SEQ ID NO:18 is 94% identical from amino acids 300 to 1771 to mouse ATP-binding cassette 2 transporter (GenBank ID g495259) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:18 also contains an ABC transporter domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:18 is an ABC transporter. SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, and SEQ ID NO:27 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-27 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention.

Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:28-54 or that distinguish between SEQ ID NO:28-54 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 7249756H2 is the identification number of an Incyte cDNA sequence, and PROSTMY01 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 71753989V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g7457275) which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. For example, GNN.g7160536\_000034\_002 is the identification number of a Genscan-predicted coding sequence, with g7160536 being the GenBank identification number of the sequence to which Genscan was applied. The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL180719\_00001 represents a "stitched" sequence in which 180719 is the identification number of the cluster of sequences to which the algorithm was applied, and 00001 is the number of the prediction generated by the algorithm. (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon-stretching" algorithm. For example, FL7472537\_g5815493\_g7406950 is the identification number of a "stretched" sequence, with 7472537 being the Incyte project identification number, g5815493 being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, and g7406950 being the GenBank identification number of the nearest GenBank protein homolog. (See Example V.) In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide

sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

5 The invention also encompasses TRICH variants. A preferred TRICH variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the TRICH amino acid sequence, and which contains at least one functional or structural characteristic of TRICH.

10 The invention also encompasses polynucleotides which encode TRICH. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:28-54, which encodes TRICH. The polynucleotide sequences of SEQ ID NO:28-54, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

15 The invention also encompasses a variant of a polynucleotide sequence encoding TRICH. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding TRICH. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:28-  
20 54 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:28-54. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of TRICH.

25 It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding TRICH, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the  
30 polynucleotide sequence of naturally occurring TRICH, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode TRICH and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring TRICH under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding TRICH or

its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding TRICH and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode TRICH and TRICH derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding TRICH or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:28-54 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding TRICH may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed,

restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060.) Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode TRICH may be cloned in recombinant DNA molecules that direct expression of TRICH, or

fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express TRICH.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter TRICH-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of TRICH, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding TRICH may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, TRICH itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of TRICH, or any part thereof, may be altered during direct synthesis and/or combined with

sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.)

5 The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, *supra*, pp. 28-53.)

In order to express a biologically active TRICH, the nucleotide sequences encoding TRICH or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in  
10 a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding TRICH. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding TRICH. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where  
15 sequences encoding TRICH and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both  
20 natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding TRICH and appropriate transcriptional and translational control  
25 elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences  
30 encoding TRICH. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or

animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding TRICH. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding TRICH can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding TRICH into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of TRICH are needed, e.g. for the production of antibodies, vectors which direct high level expression of TRICH may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of TRICH. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of TRICH. Transcription of sequences encoding TRICH may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J.



6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated  
5 transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding TRICH may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader  
10 sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses TRICH in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

15 Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

20 For long term production of recombinant proteins in mammalian systems, stable expression of TRICH in cell lines is preferred. For example, sequences encoding TRICH can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before  
25 being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These  
30 include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *ap<sup>r</sup>* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat*

confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding TRICH is inserted within a marker gene sequence, transformed cells containing sequences encoding TRICH can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding TRICH under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding TRICH and that express TRICH may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of TRICH using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on TRICH is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding TRICH include

oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide.

Alternatively, the sequences encoding TRICH, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase  
5 such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

10 Host cells transformed with nucleotide sequences encoding TRICH may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode TRICH may be designed to contain signal sequences which direct  
15 secretion of TRICH through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the  
20 protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

25 In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding TRICH may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric TRICH protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of TRICH activity. Heterologous protein and  
30 peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins,

respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the TRICH encoding sequence and the heterologous protein sequence, so that

5 TRICH may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled TRICH may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These  
10 systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

TRICH of the present invention or fragments thereof may be used to screen for compounds that specifically bind to TRICH. At least one and up to a plurality of test compounds may be screened  
15 for specific binding to TRICH. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of TRICH, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):  
20 Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which TRICH binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express TRICH, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E.  
25 coli. Cells expressing TRICH or cell membrane fractions which contain TRICH are then contacted with a test compound and binding, stimulation, or inhibition of activity of either TRICH or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the  
30 assay may comprise the steps of combining at least one test compound with TRICH, either in solution or affixed to a solid support, and detecting the binding of TRICH to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

TRICH of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of TRICH. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for TRICH activity, wherein TRICH is combined with at least one test compound, and the activity of TRICH in the presence of a test compound is compared with the activity of TRICH in the absence of the test compound. A change in the activity of TRICH in the presence of the test compound is indicative of a compound that modulates the activity of TRICH. Alternatively, a test compound is combined with an in vitro or cell-free system comprising TRICH under conditions suitable for TRICH activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of TRICH may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding TRICH or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding TRICH may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding TRICH can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region

of a polynucleotide encoding TRICH is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress TRICH, e.g., by secreting TRICH in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

### THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of TRICH and transporters and ion channels. In addition, the expression of TRICH is closely associated with normal tissues such as liver, ileum, skin, brain, dorsal root ganglion, breast, kidney, lung, pancreas, small intestine, seminal vesicle and placental tissues; normal cells such as promonocytes and bone marrow cells; and tumor tissues such as prostate, frontal lobe, pancreatic, ileal, colon and spleen tumor tissues. Therefore, TRICH appears to play a role in transport, neurological, muscle, immunological, and cell proliferative disorders. In the treatment of disorders associated with increased TRICH expression or activity, it is desirable to decrease the expression or activity of TRICH. In the treatment of disorders associated with decreased TRICH expression or activity, it is desirable to increase the expression or activity of TRICH.

Therefore, in one embodiment, TRICH or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH. Examples of such disorders include, but are not limited to, a transport disorder such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral neoplasms, prostate cancer, cardiac disorders associated with transport, e.g., angina, bradyarrhythmia, tachyarrhythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis, neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia, and other disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's

disease, Addison's disease, glucose-galactose malabsorption syndrome, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn syndrome, von Gierke disease, cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, ethanol myopathy, angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's syndrome, hypertension, hypoglycemia, myocardial infarction, migraine, pheochromocytoma, and myopathies including encephalopathy, epilepsy, Kearns-Sayre syndrome, lactic acidosis, myoclonic disorder, ophthalmoplegia, and acid maltase deficiency (AMD, also known as Pompe's disease); an immunological disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's

thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, 5 Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, 10 lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing TRICH or a fragment or derivative 15 thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified TRICH in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH including, but not limited to, 20 those provided above.

In still another embodiment, an agonist which modulates the activity of TRICH may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH including, but not limited to, those listed above.

In a further embodiment, an antagonist of TRICH may be administered to a subject to treat or 25 prevent a disorder associated with increased expression or activity of TRICH. Examples of such disorders include, but are not limited to, those transport, neurological, muscle, immunological, and cell proliferative disorders described above. In one aspect, an antibody which specifically binds TRICH may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express TRICH.

30 In an additional embodiment, a vector expressing the complement of the polynucleotide encoding TRICH may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of TRICH including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate



therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of TRICH may be produced using methods which are generally known in the art. In particular, purified TRICH may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind TRICH. Antibodies to TRICH may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with TRICH or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to TRICH have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of TRICH amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to TRICH may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc.

Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce TRICH-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be  
5 generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter,  
10 G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for TRICH may also be generated. For example, such fragments include, but are not limited to,  $F(ab')_2$  fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the  $F(ab')_2$  fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and  
15 easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such  
20 immunoassays typically involve the measurement of complex formation between TRICH and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering TRICH epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques  
25 may be used to assess the affinity of antibodies for TRICH. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of TRICH-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple TRICH epitopes, represents the average affinity, or avidity, of the antibodies for TRICH.  
30 The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular TRICH epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the TRICH-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar

procedures which ultimately require dissociation of TRICH, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

5           The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of TRICH-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for  
10 antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

          In another embodiment of the invention, the polynucleotides encoding TRICH, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules  
15 (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding TRICH. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding TRICH. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

20           In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995)  
25 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et  
30 al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

          In another embodiment of the invention, polynucleotides encoding TRICH may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-

linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in TRICH expression or regulation causes disease, the expression of TRICH from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in TRICH are treated by constructing mammalian expression vectors encoding TRICH and introducing these vectors by mechanical means into TRICH-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of TRICH include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). TRICH may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the

FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding TRICH from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to TRICH expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding TRICH under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) *J. Virol.* 61:1647-1650; Bender, M.A. et al. (1987) *J. Virol.* 61:1639-1646; Adam, M.A. and A.D. Miller (1988) *J. Virol.* 62:3802-3806; Dull, T. et al. (1998) *J. Virol.* 72:8463-8471; Zufferey, R. et al. (1998) *J. Virol.* 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding TRICH to cells which have one or more genetic abnormalities with respect to the expression of TRICH. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to

be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding TRICH to target cells which have one or more genetic abnormalities with respect to the expression of TRICH. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing TRICH to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding TRICH to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for TRICH into the alphavirus genome in place of the capsid-coding region results in the production of a large number of

TRICH-coding RNAs and the synthesis of high levels of TRICH in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of TRICH into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, *Molecular and Immunologic Approaches*, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding TRICH.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis.

Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding TRICH. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell  
5 lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be  
10 extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding TRICH. Compounds  
15 which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of  
20 polynucleotide expression. Thus, in the treatment of disorders associated with increased TRICH expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding TRICH may be therapeutically useful, and in the treatment of disorders associated with decreased TRICH expression or activity, a compound which specifically promotes expression of the polynucleotide encoding TRICH may be therapeutically useful.

25 At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound  
30 based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding TRICH is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding TRICH are assayed



by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding TRICH. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of TRICH, antibodies to TRICH, and mimetics, agonists, antagonists, or inhibitors of TRICH.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form.

These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising TRICH or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, TRICH or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example TRICH or fragments thereof, antibodies of TRICH, and agonists, antagonists or inhibitors of TRICH, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu\text{g}$  to 100,000  $\mu\text{g}$ , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind TRICH may be used for the diagnosis of disorders characterized by expression of TRICH, or in assays to monitor patients being treated with TRICH or agonists, antagonists, or inhibitors of TRICH. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for TRICH include methods which utilize the antibody and a label to detect TRICH in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring TRICH, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of TRICH expression. Normal or standard values for TRICH expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to TRICH under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of TRICH expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding TRICH may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of TRICH may be correlated

with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of TRICH, and to monitor regulation of TRICH levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding TRICH or closely related molecules may be used to identify nucleic acid sequences which encode TRICH. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding TRICH, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the TRICH encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:28-54 or from genomic sequences including promoters, enhancers, and introns of the TRICH gene.

Means for producing specific hybridization probes for DNAs encoding TRICH include the cloning of polynucleotide sequences encoding TRICH or TRICH derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding TRICH may be used for the diagnosis of disorders associated with expression of TRICH. Examples of such disorders include, but are not limited to, a transport disorder such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral neoplasms, prostate cancer, cardiac disorders associated with transport, e.g., angina, bradyarrhythmia, tachyarrhythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis, neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia, and other disorders associated with transport, e.g., neurofibromatosis, postherpetic

neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn syndrome, von Gierke disease, cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease; a 5 neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other 10 demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal 15 syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including 20 mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline 25 myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, ethanol myopathy, angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's syndrome, hypertension, hypoglycemia, myocardial infarction, migraine, pheochromocytoma, and myopathies including encephalopathy, epilepsy, Kearns-Sayre syndrome, lactic acidosis, myoclonic disorder, 30 ophthalmoplegia, and acid maltase deficiency (AMD, also known as Pompe's disease); an immunological disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact

dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding TRICH may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered TRICH expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding TRICH may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding TRICH may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding TRICH in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of TRICH, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding TRICH, under conditions suitable for hybridization or

amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding TRICH may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding TRICH, or a fragment of a polynucleotide complementary to the polynucleotide encoding TRICH, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding TRICH may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding TRICH are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as

DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of TRICH include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, TRICH, fragments of TRICH, or antibodies specific for TRICH may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number



5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an

untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global  
5 pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is  
10 achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot  
15 is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass  
20 spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for TRICH to quantify  
25 the levels of TRICH expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoz, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol-  
30 or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson,

N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

5 In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the  
10 test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated  
15 with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g.,  
20 Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach,  
25 M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding TRICH may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among  
30 members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat.

Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP).

5 (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding TRICH on a  
10 physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse,  
15 may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation.  
20 (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, TRICH, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug  
25 screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between TRICH and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT  
30 application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with TRICH, or fragments thereof, and washed. Bound TRICH is then detected by methods well known in the art. Purified TRICH can also be coated directly onto plates for use in the aforementioned drug screening techniques.

Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a

solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding TRICH specifically compete with a test compound for binding TRICH. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with TRICH.

In additional embodiments, the nucleotide sequences which encode TRICH may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/208,424, U.S. Ser. No. 60/209,001, U.S. Ser. No. 60/210,588, U.S. Ser. No. 60/212,335, U.S. Ser. No. 60/213,747, and U.S. Ser. No. 60/215,391, are hereby expressly incorporated by reference.

## EXAMPLES

### I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA

libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

## 15 II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

## III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared

using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the  
5 ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

10 The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and  
15 BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences.

20 Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive  
25 the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide  
30 sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:28-54. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

#### **IV. Identification and Editing of Coding Sequences from Genomic DNA**

Putative transporters and ion channels were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode transporters and ion channels, the encoded polypeptides were analyzed by querying against PFAM models for transporters and ion channels. Potential transporters and ion channels were also identified by homology to Incyte cDNA sequences that had been annotated as transporters and ion channels. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or



unedited Genscan-predicted coding sequences.

## V. Assembly of Genomic Sequence Data with cDNA Sequence Data

### "Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene  
5 identification program described in Example IV. Partial cDNAs assembled as described in Example  
III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan  
exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm  
based on graph theory and dynamic programming to integrate cDNA and genomic information,  
generating possible splice variants that were subsequently confirmed, edited, or extended to create a  
10 full length sequence. Sequence intervals in which the entire length of the interval was present on  
more than one sequence in the cluster were identified, and intervals thus identified were considered to  
be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic  
sequences, then all three intervals were considered to be equivalent. This process allows unrelated  
but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals  
15 thus identified were then "stitched" together by the stitching algorithm in the order that they appear  
along their parent sequences to generate the longest possible sequence, as well as sequence variants.  
Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or  
genomic sequence to genomic sequence) were given preference over linkages which change parent  
type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared  
20 by BLAST analysis to the genpept and gbprl public databases. Incorrect exons predicted by Genscan  
were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended  
with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

### "Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST  
25 analysis. First, partial cDNAs assembled as described in Example III were queried against public  
databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases  
using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST  
analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in  
Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs  
30 (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions  
may occur in the chimeric protein with respect to the original GenBank protein homolog. The  
GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous  
genomic sequences from the public human genome databases. Partial DNA sequences were  
therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant

stretched sequences were examined to determine whether it contained a complete gene.

## VI. Chromosomal Mapping of TRICH Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:28-54 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other  
5 implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:28-54 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for  
10 Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between  
15 chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site  
20 (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:8 was mapped to chromosome 12 within the interval from 137.50 to 160.90 centiMorgans.

## VII. Analysis of Polynucleotide Expression

25 Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related  
30 molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum} \{ \text{length}(\text{Seq. 1}), \text{length}(\text{Seq. 2}) \}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding TRICH are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding TRICH. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

### VIII. Extension of TRICH Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One

primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ , and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on

antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

#### 15 IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:28-54 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 µCi of [γ-<sup>32</sup>P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10<sup>7</sup> counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

#### X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing

photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Skena (1999), supra.).

Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Skena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

#### **Tissue or Cell Sample Preparation**

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated

using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14  $\mu$ l 5X SSC/0.2% SDS.

#### Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5  $\mu$ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1  $\mu$ l of the array element DNA, at an average concentration of 100 ng/ $\mu$ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

#### Hybridization

Hybridization reactions contain 9  $\mu$ l of sample mixture consisting of 0.2  $\mu$ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140  $\mu$ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

#### Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide  
5 containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477,  
10 Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

15 The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore,  
20 are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC  
25 computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

30 A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).



## XI. Complementary Polynucleotides

Sequences complementary to the TRICH-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring TRICH. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of TRICH. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the TRICH-encoding transcript.

## XII. Expression of TRICH

Expression and purification of TRICH is achieved using bacterial or virus-based expression systems. For expression of TRICH in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express TRICH upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of TRICH in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding TRICH by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, TRICH is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from TRICH at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification

using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10 and 16). Purified TRICH obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII, where applicable.

### XIII. Functional Assays

TRICH function is assessed by expressing the sequences encoding TRICH at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of TRICH on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding TRICH and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding TRICH and other genes of interest can be analyzed by northern

analysis or microarray techniques.

#### **XIV. Production of TRICH Specific Antibodies**

TRICH substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the TRICH amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, *supra*, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, *supra*.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-TRICH activity by, for example, binding the peptide or TRICH to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### **XV. Purification of Naturally Occurring TRICH Using Specific Antibodies**

Naturally occurring or recombinant TRICH is substantially purified by immunoaffinity chromatography using antibodies specific for TRICH. An immunoaffinity column is constructed by covalently coupling anti-TRICH antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing TRICH are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of TRICH (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/TRICH binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and TRICH is collected.

#### **XVI. Identification of Molecules Which Interact with TRICH**

Molecules which interact with TRICH may include transporter substrates, agonists or antagonists, modulatory proteins such as G $\beta\gamma$  proteins (Reimann, *supra*) or proteins involved in TRICH localization or clustering such as MAGUKs (Craven, *supra*). TRICH, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973))

Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled TRICH, washed, and any wells with labeled TRICH complex are assayed. Data obtained using different concentrations of TRICH are used to calculate values for the number, affinity, and association of TRICH with the candidate molecules.

5 Alternatively, proteins that interact with TRICH are isolated using the yeast 2-hybrid system (Fields, S. and O. Song (1989) Nature 340:245-246). TRICH, or fragments thereof, are expressed as fusion proteins with the DNA binding domain of Gal4 or lexA, and potential interacting proteins are expressed as fusion proteins with an activation domain. Interactions between the TRICH fusion protein and the TRICH interacting proteins (fusion proteins with an activation domain) reconstitute a  
10 transactivation function that is observed by expression of a reporter gene. Yeast 2-hybrid systems are commercially available, and methods for use of the yeast 2-hybrid system with ion channel proteins are discussed in Niethammer, M. and M. Sheng (1998, Meth. Enzymol. 293:104-122).

TRICH may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions  
15 between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

Potential TRICH agonists or antagonists may be tested for activation or inhibition of TRICH ion channel activity using the assays described in section XVIII.

#### **XVII. Demonstration of TRICH Activity**

20 Ion channel activity of TRICH is demonstrated using an electrophysiological assay for ion conductance. TRICH can be expressed by transforming a mammalian cell line such as COS7, HeLa or CHO with a eukaryotic expression vector encoding TRICH. Eukaryotic expression vectors are commercially available, and the techniques to introduce them into cells are well known to those skilled in the art. A second plasmid which expresses any one of a number of marker genes, such as  $\beta$ -  
25 galactosidase, is co-transformed into the cells to allow rapid identification of those cells which have taken up and expressed the foreign DNA. The cells are incubated for 48-72 hours after transformation under conditions appropriate for the cell line to allow expression and accumulation of TRICH and  $\beta$ -galactosidase.

Transformed cells expressing  $\beta$ -galactosidase are stained blue when a suitable colorimetric  
30 substrate is added to the culture media under conditions that are well known in the art. Stained cells are tested for differences in membrane conductance by electrophysiological techniques that are well known in the art. Untransformed cells, and/or cells transformed with either vector sequences alone or  $\beta$ -galactosidase sequences alone, are used as controls and tested in parallel. Cells expressing TRICH will have higher anion or cation conductance relative to control cells. The contribution of TRICH to

conductance can be confirmed by incubating the cells using antibodies specific for TRICH. The antibodies will bind to the extracellular side of TRICH, thereby blocking the pore in the ion channel, and the associated conductance.

Alternatively, ion channel activity of TRICH is measured as current flow across a TRICH-containing Xenopus laevis oocyte membrane using the two-electrode voltage-clamp technique (Ishi et al., supra; Jegla, T. and L. Salkoff (1997) J. Neurosci. 17:32-44). TRICH is subcloned into an appropriate Xenopus oocyte expression vector, such as pBF, and 0.5-5 ng of mRNA is injected into mature stage IV oocytes. Injected oocytes are incubated at 18°C for 1-5 days. Inside-out macropatches are excised into an intracellular solution containing 116 mM K-gluconate, 4 mM KCl, and 10 mM Hepes (pH 7.2). The intracellular solution is supplemented with varying concentrations of the TRICH mediator, such as cAMP, cGMP, or Ca<sup>2+</sup> (in the form of CaCl<sub>2</sub>), where appropriate. Electrode resistance is set at 2-5 MΩ and electrodes are filled with the intracellular solution lacking mediator. Experiments are performed at room temperature from a holding potential of 0 mV. Voltage ramps (2.5 s) from -100 to 100 mV are acquired at a sampling frequency of 500 Hz. Current measured is proportional to the activity of TRICH in the assay.

In particular the activity of TRICH-10 is measured as cation conductance in the presence of heat, the activity of TRICH-12 is measured as anion conductance in the presence of GABA, the activity of TRICH-13 is measured as Na<sup>+</sup> conductance, the activity of TRICH-21 is measured as voltage-gated Cl<sup>-</sup> conductance, the activity of TRICH-22 is measured as Ca<sup>2+</sup> conductance, the activity of TRICH-24 is measured as voltage-gated Ca<sup>2+</sup> conductance, the activity of TRICH-26 is measured as K<sup>+</sup> conductance in the presence of cyclic nucleotides, and the activity of TRICH-27 is measured as Cl<sup>-</sup> conductance.

Transport activity of TRICH is assayed by measuring uptake of labeled substrates into Xenopus laevis oocytes. Oocytes at stages V and VI are injected with TRICH mRNA (10 ng per oocyte) and incubated for 3 days at 18°C in OR2 medium (82.5mM NaCl, 2.5 mM KCl, 1mM CaCl<sub>2</sub>, 1mM MgCl<sub>2</sub>, 1mM Na<sub>2</sub>HPO<sub>4</sub>, 5 mM Hepes, 3.8 mM NaOH, 50μg/ml gentamycin, pH 7.8) to allow expression of TRICH. Oocytes are then transferred to standard uptake medium (100mM NaCl, 2 mM KCl, 1mM CaCl<sub>2</sub>, 1mM MgCl<sub>2</sub>, 10 mM Hepes/Tris pH 7.5). Uptake of various substrates (e.g., amino acids, sugars, drugs, ions, and neurotransmitters) is initiated by adding labeled substrate (e.g., radiolabeled with <sup>3</sup>H, fluorescently labeled with rhodamine, etc.) to the oocytes. After incubating for 30 minutes, uptake is terminated by washing the oocytes three times in Na<sup>+</sup>-free medium, measuring the incorporated label, and comparing with controls. TRICH activity is proportional to the level of internalized labeled substrate. In particular, test substrates include organic cations for TRICH-9, carnitine and acylcarnitine for TRICH-11, galactose and other sugars for TRICH-14, glucose for

TRICH-15, monocarboxylate for TRICH-16, cations for TRICH-17, estramustine and related drugs for TRICH-18, amino acids for TRICH-19, glucose for TRICH-20, sugars for TRICH-23, and glucose or fructose for TRICH-25.

In the alternative, TRICH transport activity can be demonstrated through the use of a ligand mixing assay that is used to measure transport from early to late endosomal compartments in X. laevis oocytes. Ovaries are dissected from adult female X. laevis, and oocytes are isolated. (Mukhopadhyay A. et al. (1997) J. Cell. Biol. 136(6): 1227-1237). Oocytes are pulsed with 2mg/ml avidin for 5hrs at 18° C, washed, then incubated for 16 hrs to allow avidin to transport to a late compartment. The oocytes are then incubated with 1mg/ml biotin-horseradish peroxidase (HRP) for 30 minutes at 18° C to label early endocytic compartments. Varying amounts of TRICH are injected into the oocytes, and the oocytes are incubated at 18° C. Oocytes are collected at several time points after TRICH injection, washed, and lysed in 100µl of phosphate-buffered saline containing 0.3% Triton X-100, 0.2% methylbenzethorium chloride, and 400 µg/ml of BSA-biotin as a scavenger. Finally, the lysates are centrifuged for 30 seconds in a microfuge, and the avidin-biotin complexes are immunoprecipitated using anti-avidin antibody-coated plates by incubation at 4 °C overnight. The plates are washed at least 5 times to remove unbound proteins. Transport from the early endosomes to the late compartments is quantified by measuring the amount of immunoprecipitated HRP; increased transport due to TRICH is quantitated by comparison with control oocytes. Potential inhibitors of proton-dependent histidine transport such as dipeptides and tripeptides can subsequently be tested in the expression system described above (Yamashita, T. et al. (1997) J. Cell. Biol. 136(6): 1227-1237).

ATPase activity associated with TRICH can be measured by hydrolysis of radiolabeled ATP- $[\gamma\text{-}^{32}\text{P}]$ , separation of the hydrolysis products by chromatographic methods, and quantitation of the recovered  $^{32}\text{P}$  using a scintillation counter. The reaction mixture contains ATP- $[\gamma\text{-}^{32}\text{P}]$  and varying amounts of TRICH in a suitable buffer incubated at 37°C for a suitable period of time. The reaction is terminated by acid precipitation with trichloroacetic acid and then neutralized with base, and an aliquot of the reaction mixture is subjected to membrane or filter paper-based chromatography to separate the reaction products. The amount of  $^{32}\text{P}$  liberated is counted in a scintillation counter. The amount of radioactivity recovered is proportional to the ATPase activity of TRICH in the assay.

### **XVIII. Identification of TRICH Agonists and Antagonists**

TRICH is expressed in a eukaryotic cell line such as CHO (Chinese Hamster Ovary) or HEK (Human Embryonic Kidney) 293. Ion channel activity of the transformed cells is measured in the presence and absence of candidate agonists or antagonists. Ion channel activity is assayed using patch clamp methods well known in the art or as described in Example XVII. Alternatively, ion

channel activity is assayed using fluorescent techniques that measure ion flux across the cell membrane (Velicelebi, G. et al. (1999) Meth. Enzymol. 294:20-47; West, M.R. and C.R. Molloy (1996) Anal. Biochem. 241:51-58). These assays may be adapted for high-throughput screening using microplates. Changes in internal ion concentration are measured using fluorescent dyes such as the

5  $\text{Ca}^{2+}$  indicator Fluo-4 AM, sodium-sensitive dyes such as SBFI and sodium green, or the  $\text{Cl}^-$  indicator MQAE (all available from Molecular Probes) in combination with the FLIPR fluorimetric plate reading system (Molecular Devices). In a more generic version of this assay, changes in membrane potential caused by ionic flux across the plasma membrane are measured using oxonol dyes such as DiBAC<sub>4</sub> (Molecular Probes). DiBAC<sub>4</sub> equilibrates between the extracellular solution and cellular sites

10 according to the cellular membrane potential. The dye's fluorescence intensity is 20-fold greater when bound to hydrophobic intracellular sites, allowing detection of DiBAC<sub>4</sub> entry into the cell (Gonzalez, J.E. and P.A. Negulescu (1998) Curr. Opin. Biotechnol. 9:624-631). Candidate agonists or antagonists may be selected from known ion channel agonists or antagonists, peptide libraries, or combinatorial chemical libraries.

15

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be

20 understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
7475353	1	7475353CD1	28	7475353CB1
3107278	2	3107278CD1	29	3107278CB1
7473394	3	7473394CD1	30	7473394CB1
7473900	4	7473900CD1	31	7473900CB1
7475045	5	7475045CD1	32	7475045CB1
7475611	6	7475611CD1	33	7475611CB1
7475617	7	7475617CD1	34	7475617CB1
7473314	8	7473314CD1	35	7473314CB1
70356714	9	70356714CD1	36	70356714CB1
7611491	10	7611491CD1	37	7611491CB1
171968	11	171968CD1	38	171968CB1
257274	12	257274CD1	39	257274CB1
6355991	13	6355991CD1	40	6355991CB1
70035348	14	70035348CD1	41	70035348CB1
7472539	15	7472539CD1	42	7472539CB1
817477	16	817477CD1	43	817477CB1
1442166	17	1442166CD1	44	1442166CB1
2311751	18	2311751CD1	45	2311751CB1
7472537	19	7472537CD1	46	7472537CB1
7472546	20	7472546CD1	47	7472546CB1
7474202	21	7474202CD1	48	7474202CB1
7476280	22	7476280CD1	49	7476280CB1
1713377	23	1713377CD1	50	1713377CB1
5842557	24	5842557CD1	51	5842557CB1
7476643	25	7476643CD1	52	7476643CB1
7611651	26	7611651CD1	53	7611651CB1
2522075	27	2522075CD1	54	2522075CB1



Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	7475353CD1	g2982567	0	ABC transporter [Rattus norvegicus] (Hirsch-Ernst, K. I. et al. (1998) Molecular cDNA cloning and tissue distribution of mRNA encoding a novel ATP-binding cassette (ABC) half-transporter. Biochem. Biophys. Res. Commun. 249:151-155.)
2	3107278CD1	g6010763	5.00E-180	ion transporter protein [Rattus norvegicus]
3	7473394CD1	g12724309	0	sugar ABC transporter ATP binding protein [Lactococcus lactis subsp. lactis] (Bolotin, A. et al. (2001) The Complete Genome Sequence of the Lactic Acid Bacterium Lactococcus lactis ssp. lactis IL1403. Genome Res. 11:731-753.)
4	7473900CD1	g4980593	2.20E-131	sugar ABC transporter, ATP-binding protein [Thermotoga maritima]
		g5565862	8.10E-141	ornithine transporter [Homo sapiens] (Camacho, J.A. et al. (1999) Hyperornithinaemia- hyperammonaemia-homocitrullinuria syndrome is caused by mutations in a gene encoding a mitochondrial ornithine transporter. Nat. Genet. 22:151-158.)
5	7475045CD1	g5701943	3.50E-44	mitochondrial oxaloacetate transport protein [Saccharomyces cerevisiae] (Palmieri, L. et al. (1999) Identification of the yeast mitochondrial transporter for oxaloacetate and sulfate. J. Biol. Chem. 274:22184-22190.)
6	7475611CD1	g2808786	3.10E-61	cobalt transport system ATP binding protein [Streptomyces coelicolor]
7	7475617CD1	g2944233	1.20E-239	sodium-hydrogen exchanger 6 [Homo sapiens] (Numata, M. et al. (1998) Identification of a mitochondrial Na <sup>+</sup> /H <sup>+</sup> exchanger. J. Biol. Chem. 273:6951-6959.)

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
8	7473314CD1	g2208839	1.80E-262	peptide/histidine transporter [Rattus norvegicus] (Yamashita, T. et al. (1997) Cloning and functional expression of a brain peptide/histidine transporter. J. Biol. Chem. 272:10205-10211.)
9	70356714CD1	g7707622	0	organic anion transporter 4 [Homo sapiens] (Cha, S. H. et al. (2000) Molecular cloning and characterization of multispecific organic anion transporter 4 expressed in the placenta. J. Biol. Chem. 275:4507-4512.)
10	7611491CD1	g2696709	5.00E-141	RST (Renal specific transporter) [Mus musculus] (Mori, K. et al. (1997) Kidney-specific expression of a novel mouse organic cation transporter-like protein. FEBS Lett. 417:371-374.)
		g11055322	0	vanilloid receptor-related osmotically activated channel [Homo sapiens] (Liedtke, W. et al. (2000) Vanilloid receptor-related osmotically activated channel (VR-OAC), a candidate vertebrate osmoreceptor. Cell 103:525-535.)
		g2570933	6.90E-135	vanilloid receptor subtype 1 [Rattus norvegicus] (Caterina, M.J. et al. (1997) The capsaicin receptor: a heat-activated ion channel in the pain pathway. Nature 389:816-824.)
11	171968CD1	g13027346	2.00E-33	putative carnitine/acylcarnitine translocase [Oryza sativa]
		g4239974	6.90E-25	mCAC (mitochondrial carnitine/acylcarnitine transporter) [Mus musculus]
12	257274CD1	g292040	1.30E-39	GABA-alpha receptor beta-3 subunit [Homo sapiens]

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
13	6355991CD1	g12642270	0	voltage-gated sodium channel alpha subunit SCN1A [Homo sapiens]
		g1041089	0	Na <sup>+</sup> channel [Rattus norvegicus]
14	70035348CD1	g1789312	1.30E-53	(Noda, M. and Numa, S. (1987) Structure and function of sodium channel. J. Recept. Res. 7:467-497.)
15	7472539CD1	g9588428	0	galactose-proton symport of transport system [Escherichia coli]
				dJ1024N4.1 (novel Sodium:solute symporter family member similar to SLC5A1 (SGLT1))
				[Homo sapiens]
16	817477CD1	g286259	8.90E-174	Sodium/glucose cotransporter [Rattus norvegicus]
		g6093322	8.00E-63	monocarboxylate transporter MCT3 [Homo sapiens]
				(Yoon, H., et al. (1999) Cloning of the human monocarboxylate transporter MCT3 gene: localization to chromosome 22q12.3-q13.2. Genomics 60:366-370.)
17	1442166CD1	g12248394	0	cation-transporting ATPase [Mus musculus]
18	2311751CD1	g495259	0	abc2 [Mus musculus]
				(Laing N.M. et al. (1998) Amplification of the ATP- binding cassette 2 transporter gene is functionally linked with enhanced efflux of estramustine in ovarian carcinoma cells. Cancer Res. 58:1332-1337.)
19	7472537CD1	g7406950	2.00E-137	N system amino acids transporter NAT-1 [Mus musculus]
				(Gu, S. et al. (2000) Identification and characterization of an amino acid transporter expressed differentially in liver. Proc. Natl. Acad. Sci. U.S.A. 97:3230-3235.)

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
20	7472546CD1	g9588428 g338055	0 2.70E-197	dJ1024N4.1 (novel Sodium:solute symporter family member similar to SLC5A1 (SGLT1)) [Homo sapiens] Na+/glucose cotransporter [Homo sapiens] (Hediger, M.A. et al. (1989) Homology of the human intestinal Na+/glucose and Escherichia coli Na+/proline cotransporters. Proc. Natl. Acad. Sci. U.S.A. 86:5748-5752.) ClC chloride channel ClC-K2 (human, kidney) [Homo sapiens] (Takeuchi, Y. et al. (1995) Cloning, tissue distribution, and intrarenal localization of ClC chloride channels in human kidney. Kidney Int. 48:1497-1503.)
21	7474202CD1	g1217689	0	TRP2 (transient receptor potential) [Rattus norvegicus] (Liman, E.R. et al. (1999) TRP2: a candidate transduction channel for mammalian pheromone sensory signaling. Proc. Natl. Acad. Sci. U.S.A. 96:5791-5796.) Similarity to Yeast low-affinity glucose transporter HXT4 [Caenorhabditis elegans] voltage-gated calcium channel [Rattus norvegicus] (Ishibashi, K. et al. (2000) Molecular cloning of a novel form (Two-repeat) protein related to voltage-gated sodium and calcium channels. Biochem. Biophys. Res. Commun. 270:370-376.)
22	7476280CD1	g4877836	0	
23	1713377CD1	g3874275	2.20E-76	
24	5842557CD1	g4586963	1.10E-23	

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
25	7476643CD1	g9230651	0	facilitative glucose transporter family member GLUT9 [Homo sapiens] (Phay, J.E. et al. (2000) Cloning and expression analysis of a novel member of the facilitative glucose transporter family, SLC2A9 (GLUT9). Genomics 66:217-220.)
		g183298	3.70E-113	GLUT5 protein [Homo sapiens] (Kayano, T. et al. (1990) Human facilitative glucose transporters. Isolation, functional characterization, and gene localization of cDNAs encoding an isoform (GLUT5) expressed in small intestine, kidney, muscle, and adipose tissue and an unusual glucose transporter pseudogene-like sequence (GLUT6). J. Biol. Chem. 265:13276-13282.)
26	7611651CD1	g2745729	0	potassium channel [Rattus norvegicus] (Shi, W. et al. (1997) Identification of two nervous system-specific members of the erg potassium channel gene family. J. Neurosci. 17:9423-9432.)
27	2522075CD1	g7592636	1.60E-183	Parchorin [Oryctolagus cuniculus] (Nishizawa, T. et al. (2000) Molecular cloning and characterization of a novel chloride intracellular channel-related protein, parchorin, expressed in water- secreting cells. J. Biol. Chem. 275:11164-11173.)

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7475353CD1	842	S111 S34 S342 S357 S598 S661 S728 S755 T238 T294 T341 T394 T462 T605 T679 T759 T831 T97	N447 N498 N677 N775	ABC TRANSPORTERS FAMILY DM00008 Q02592 583-793:R589-G800 ABC TRANSPORTER PDI30117: M1-L263 ABC transporters family BL00211A:L621-I632 BL00211B: L727-D758 ABC transporters family signature atp_bind_transport.prf:A708-D758 Transmembrane domain: F186-G203, Y386-S406 ABC transporter transmembrane region. ABC_membrane:V265-L544 ABC transporter ABC_tran:G616-G800 Abc_Transporter L727-I741 Atp_Gtp_A G623-S630	BLAST_DOMO BLAST_PRODUM BLIMPS_BLOCKS ProfileScan HMMER HMMER_PFAM HMMER_PFAM Motifs Motifs

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2	3107278CD1	461	S141 S153 S185 S306 S406 T162 T429 T448 T60	N404 N54	Signal peptide: M1-S52  Transmembrane domains: I39-F58; S59-T75; V119-F138; G270-N290; V341-T360; F373-L392  Sugar (and other) transporter domain: S5-E409 (Score = -64.7; E-value = 1.5e-4) Sugar transport proteins signature BL00216: F58-M107  Sugar transport motif: T22-S38	SPSCAN  HMMER  HMMER_PPFAM  BLIMPS_BLOCKS  MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	7473394CD1	485	S236 S462 S52 T10 T166 T191 T239 T316 T324 T345 T386 T89	N3 N367 N460	Signal peptide: M1-G53  ABC transporter domain: G24-G210; G277-G471  ABC transporters family signature BL00211: L29-L40; L396-D427  ABC transporters family signature: L378-D427  ATPBINDING PUTATIVE ATPASE RIBOSE /GALACTOSE ABC TRANSPORTER PROTEIN MGLA PD035715: K241-K311  ABC TRANSPORTERS FAMILY DM00008 P47365 6-219: M1-R208; E260-I468  ABC transporter motif: L396-V410  ATP/GTP binding site (P-loop): G31-S38	SPSCAN  HMMER_PFAM  BLIMPS_BLOCKS  PROFILESCAN  BLAST_PRODOR  BLAST_DOMO  MOTIFS  MOTIFS



Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	7473900CD1	301	S143 S200 S203 S290 T136 T32 T39		Mitochondrial carrier proteins domain: Q8-M294  Mitochondrial energy transfer protein signature BL00215: L214-Q238  Mitochondrial energy transfer proteins signature: A10-G59; Q101-K163; K204-A276 PROTEIN TRANSPORT TRANSMEMBRANE REPEAT MITOCHONDRION CARRIER MEMBRANE INNER MITOCHONDRIAL ADP/ATP PD000117: Y44-S241  MITOCHONDRIAL ENERGY TRANSFER PROTEINS DM00026 S55056 202-289: L207-E288  Mitochondrial carrier protein motif: P126-L134; P229-I237	HMMEP_PPFAM  BLIMPS_BLOCKS  PROFILES CAN  BLAST_PRODOM  BLAST_DOMO  MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	7475045CD1	304	S190 S231 T225		Signal peptide: M1-A20	HMMER
					Mitochondrial carrier proteins domain: P5-A296	HMMER_PFAM
					Mitochondrial energy transfer proteins signature BL00215: V11-Q35; I258-G270	BLIMPS_BLOCKS
					Mitochondrial energy transfer proteins signature: A7-V60; W206-I258	PROFILES CAN
					PROTEIN TRANSPORT TRANSMEMBRANE REPEAT MITOCHONDRION CARRIER MEMBRANE INNER MITOCHONDRIAL ADP/ATP PD000117: D9-Y91; T100-K294	BLAST_PROD OM
					MITOCHONDRIAL ENERGY TRANSFER PROTEINS DM00026 P32332 233-312: L210-L295	BLAST_DOMO
					Mitochondrial carrier protein motif: P26-L34	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6	7475611CD1	278	S144 S2 S29 S44 S56 T260	N27 N42	ABC transporter domain: G33-G218 ABC transporters family signature BL00211: I38-L49; L143-D174 ABC transporters family signature: S124-D174 COBALT TRANSPORT SYSTEM ATP BINDING PROTEIN MEMBRANE ASSOCIATED ATPASE PD029284: D186-S269 ABC TRANSPORTERS FAMILY DM00008 Q05596 2-210: L7-L204 ABC transporter motif: L143-V157 ATP/GTP binding site (P-loop): G40-T47	HMMER_PFAM BLIMPS_BLOCKS PROFILES CAN BLAST_PRODOM BLAST_DOMO MOTIFS MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	7475617CD1	673	S13 S145 S191 S207 S493 S532 S636 S641 S642 S659 S71 S94 T101 T124 T538 T6 T605 T612 T631 T80	N348 N519 N536 N621 N92	Transmembrane domains: V19-F38; L155-I173; L275-T296; M457-T477 Sodium/hydrogen exchanger family domain: L21-V487 Na+/H+ exchanger signature PR01084: V129-F140; G143-S157; I158-T166; G203-T213 Na+/H+ exchanger isoform PR01088: E11-I35; W36-I54; Y55-Q81; E115-E128; S246-D263; A265-M284; T476-W502; G535-D553; P559-Q587; V588-D615 + TRANSPORT EXCHANGER NA PD01672: V129-M177 SODIUMHYDROGEN EXCHANGER 6 MYELOBLAST KIAA0267 PD177855: G474-E494; Y504-N672 do BETA; EXCHANGER; NA; DM02572 P48764 10-734: E11-L63; D118-R486	HMMER  HMMER_PPFAM BLIMPS_PRINTS BLIMPS_PRINTS BLIMPS_PRODOR BLAST_PRODOR BLAST_DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8	7473314CD1	576	S134 S269 S278 S289 S293 S297 S400 S423 S510 S519 S553 S573 T169 T190 T369 T572	N139 N218 N355 N435	PTR2 FAMILY PROTON/OLIGOPEPTIDE SYMPORTERS DM01990: A35-N525, P305-C529, V311-W517, A34-G524 TRANSPORTER TRANSMEMBRANE PEPTIDE OLIGOPEPTIDE PROTEIN SYMPORT ISOFORM H+/PEPTIDE COTRANSPORTER PD001550: T307-S499 PEPTIDE/HISTIDINE TRANSPORTER PD127516: F494-R575 PTR2A PEPTIDE TRANSPORTER TRANSPORT TRANSMEMBRANE PD170949: D447-S499 PTR2 family proton/oligopeptide transporter BL01022E: E464-S499, E43-L61, A73-A118, G159-V182, F194-I206 PTR2 Proton-dependent oligopeptide transport (POT) family peptide transporter signature: A102-S495 PTR2 family proton/oligopeptide symporters signature 2: F194-I206 Multicopper Oxidase signature 1: G489-V509 Transmembrane domain: L401-L421, M483-V501, I526-I551	BLAST-DOMO BLAST-PRODOM BLAST-PRODOM BLIMPS-BLOCKS HMMER-PFAM MOTIFS MOTIFS HMMER

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	70356714CD1	550	S104 S106 S164 S225 S279 S319 S326 S332 S529 T224 T428 T523 T65	N310 N353 N39 N56 N99	Transmembrane domain: I148-Y165 Sugar (and other) transporter domain: T103-L527	HMNER HMNER_PFAM
10	7611491CD1	559	S105 S110 S120 S129 S347 S376 S47 S524 S91 T114 T192 T428 T68 T83 Y99	N339 N472 N490	Transmembrane domains: A156-Y178; V203-F221; L239-Y262; A261-F280; F305-L324; P380-N400 PROTEIN OLFACTORY CHANNEL B0212.5 T09A12.3 T10B10.7 VANILLOID RECEPTOR SUBTYPE F28H7.10 PD011151: N54-P186 VANILLOID RECEPTOR SUBTYPE 1 PD137334: L440-P515	HMNER BLAST_PRODROM
11	171968CD1	181	S142 S159 T113 T64	N22	Mitochondrial carrier proteins domains: C9-E79; Y85-W180 Mitochondrial energy transfer proteins signature: A95-E146 PROTEIN TRANSPORT TRANSMEMBRANE REPEAT MITOCHONDRION CARRIER MEMBRANE INNER MITOCHONDRIAL ADP/ATP PD000117: P12-W179 MITOCHONDRIAL ENERGY TRANSFER PROTEINS DM00026 P38087 243-325: G101-W179 Mitochondrial carrier proteins motif: P12-L20; P114-M122	HMNER_PFAM PROFILESCAN BLAST_PRODROM BLAST_DOMO MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12	257274CD1	124	S27 T110	N33	Signal peptide: M1-G22	SPSCAN, HMMER
					Neurotransmitter-gated ion-channel domain: K38-M80	HMMER_PFAM
					NEUROTRANSMITTER-GATED ION-CHANNELS DM00560 S53532 15-474: R26-L84	BLAST_DOMO

Table 3 (cont.)

[illegible]



Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14	70035348CD1	538	S169 S220 S256 S264 S385 S443 S495 S535 S75 T18 T246 T403 T520	N371 N383 N396 N401	Transmembrane domains: V83-I101; C115-I134; T131-V153; Y198-F216; T345-V364 Sugar (and other) transporter domain: S43-V484 Sugar transport proteins signature BL00216: G51-S62; L133-A182 Sugar transport proteins signature: L119-I184 SUGAR TRANSPORT PROTEINS DM00135 P09830 101-452: L119-G362; V426-I487 Sugar transporter motif: G97-S113	HMMER  HMMER_PFAM  BLIMPS_BLOCKS  PROFILES SCAN  BLAST_DOMO  MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
15	7472539CD1	742	S139 S157 S22 S26 S380 S494 S634 S643 S648 S699 S712 T495 T52 T558 T711 Y583	N324 N329 N476 N664	Transmembrane domains: I121-I140; I223-I240; L261-M280; L446-A466; V505-I521; L604-T622 Sodium:solute symporter family domain: I140-G569 Sodium:solute symporter signature BL00456: A193-R222; L255-G309; P542-A551 Sodium:solute symporter family signatures: Q252-V299; D531-D592 TRANSMEMBRANE TRANSPORT PERMEASE PROTEIN SODIUM SYMPORT PROLINE COTRANSPORTER SYMPORTER GLYCOPROTEIN PD000991: V197-G569 SODIUM:SOLUTE SYMPORTER FAMILY DM00745 S59637 24-561: H117-T625 Sodium solute symporter motif: G256-A281 Transmembrane domains: V82-S109; I338-G356 Monocarboxylate transporter domain: A20-D426 do PEST; TRANSPORTER; LINKED; DM05037 P53988 1-465: P7-P191; S209-L389	HMMER  HMMER_PFAM  BLIMPS_BLOCKS  PROFILES CAN  BLAST_DOMO  MOTIFS  HMMER  HMMER_PFAM  BLAST_DOMO
16	817477CD1	426	S134 S138 S193 S74 T335	N369		

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17	1442166CD1	1197	S205 S224 S306 S328 S612 S634 S712 S740 S776 S799 S851 S929 S1127 S1152 T438 T567 T596 T603 T66 T910 T913 T961 T1190	N150 N287 N420 N502 N738 N1100	Transmembrane domains: V67-W87; F445-I464 E1-E2 ATPase domains: Q302-H393; A524-D599; E677-A880 E1-E2 ATPases phosphorylation site proteins signature BL00154: V489-G525; V527-V545; C723-F763; T859-L882 P-type cation-transporting ATPase superfamily signature PR00119: D348-E362; C531-V545; A739-D749; C862-L881 PROBABLE CALCIUMTRANSPORTING ATPASE HYDROLASE CALCIUM TRANSPORT TRANSMEMBRANE PHOSPHORYLATION MAGNESIUM ATPBINDING PD023991: D943-G1189 E1-E2 ATPASES PHOSPHORYLATION SITE DM00115 P54678 80-795: W263-G815; E806-L881 E1-E2 ATPase motif: D533-T539	HMMER HMMER_PFAM BLIMPS_BLOCKS BLIMPS_PRINTS BLAST_PRODOR BLAST_DOMO MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
18	2311751CD1	1771	S219 S275 S294 S306 S449 S454 S468 S583 S658 S667 S674 S716 S746 S762 S790 S813 S895 S939 S1531 S1701 S1317 S1494 S1580 S1627 S1668 S1755 S1022 S1154 S1359 S1371 S1397 T179 T290 T31 T393 T416 T547 T606 T648 T649 T867 T1437 T1443 T1479 T1550 T1687 T1748 T1432 T1570 T1619 Y725	N744 N832 N885 N893 N948 N1013 N1111 N1390	Transmembrane domains: V119-L138; L228-T246; V1128-F1148; M180-F1197; V1235-L1261 ABC transporter domain: N353-G533; G1416-G1597 ABC transporters family signature: D440-D490; V1502-D1553 ATPBINDING TRANSPORTER CASSETTE ABC TRANSPORT PROTEIN GLYCOPROTEIN TRANSMEMBRANE RIM ABCR PD005939: L1122-Y1306 ABC TRANSPORTERS FAMILY DM00008 P41233 839-1045:V326-H532; V1386-M1594 ABC transporter motif: L459-F473 ATP/GTP binding site (P-loop) G360-T367; G1423-T1430	HMNER  HMNER_PFAM PROFILES SCAN BLAST_PROD OM BLAST_DOMO MOTIFS MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
19	7472537CD1	474	S22 S232 S236 S287 S436 T163 T400 T433	N296 N56	Transmembrane domains: G77-V96; I175-T198; V342-F359; W377-L397; I396-I419; L443-I462  Transmembrane amino acid transporter protein domain: A72-M453  ACID AMINO PROTEIN TRANSPORTER PERMEASE TRANSMEMBRANE INTERGENIC REGION PUTATIVE PROLINE PD001875: S53-V361  TRANSPORTER PROTEIN PD138374: H327-W464	HMMER   HMMER_PFBM  BLAST_PRODUM  BLAST_PRODUM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
20	7472546CD1	752	S139 S22 S26 S390 S504 S644 S653 S658 S709 S722 T505 T52 T568 T721 Y593	N196 N334 N339 N486 N674	Transmembrane domains: I121-I140; S173-W195; I233-I250; L271-M290; L456-A476; V515-I531; L614-T632 Sodium:solute symporter family domain: Y150-G579 Sodium:solute symporter signature BL00456: Y127-G181 A203-R232 L265-G319 P552-A561 Sodium:solute symporter family signatures: Q262-V309; D541-D602 TRANSMEMBRANE TRANSPORT PERMEASE PROTEIN SODIUM SYMPORT PROLINE COTRANSPORTER SYMPORTER GLYCOPROTEIN PD000991: Y150-G579 SODIUM:SOLUTE SYMPORTER FAMILY DM00745 S59637 24-561: H117-T635 Sodium solute symporters motif: G266-A291	HMMER    HMMER_PFAM  BLIMPS_BLOCKS  PROFILES CAN  BLAST_PROD OM  BLAST_DOMO  MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
21	7474202CD1	654	S200 S226 S238 S284 S321 S328 S486 S636 T155 T17 T280 T290 T530 T540 T626 Y51	N161 N332 N520 N646	Transmembrane domains: D49-G72; F364-I382 Voltage gated chloride channels domain: M67-Q484 CBS domains: H517-Q572; C593-S645 Chloride channel signature PR00762: P84-V101; V117-P136; A174-E193; M389-G409; G432-H448; T449-P468; F487-P501 PROTEIN CHANNEL CHLORIDE TRANSMEMBRANE VOLTAGEGATED IONIC ION TRANSPORT CBS DOMAIN PD001036: Q80-L474 do CHANNEL; CHLORIDE; CLC-1; CLC-KA; DM01220 P51800 52-686: F52-R76; G77-K654	HMMER HMMER_PFAM HMMER_PFAM BLIMPS_PRINTS BLAST_PRODUM BLAST_DOMO
22	7476280CD1	886	S136 S17 S280 S341 S40 S414 S580 S599 S678 S729 S86 S880 T282 T364 T407 T491 T702 T745 T753 T782 T788 T882	N265 N582 N7	Transmembrane domains: K343-W362; S386-L405; M546-Y571; I619-T637 Transient receptor potential signature PR01097: G618-S639; F640-F653; T668-A681 CHANNEL PROTEIN CALCIUM ENTRY CAPACITATIVE IONIC TRANSMEMBRANE ION TRANSPORT TRANSIENT PD004194: L23-H499 ANK MOTIF REPEAT DM03196 P48994 13-780: D61-I567; E575-R688; E724-E751	HMMER BLIMPS_PRINTS BLAST_PRODUM BLAST_DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
23	1713377CD1	512	S109 S132 S246 S304 S330 S508 T113 T234	N257	Transmembrane domains: A49-I72; D307-L325; L451-Y470 Sugar transport proteins signature BL00216: F139-G188	HMMEER BLIMPS_BLOCKS
24	5842557CD1	475	S115 S262 S284 S389 S97 T462	N334 N341	Transmembrane domains: A12-Y35; Y155-L178; I194-L220; V226-Y246; A304-F324 Ion transport protein domain: L151-I416	HMMEER HMMEER_PFAM



Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosyla- tion Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
25	7476643CD1	537	S193 S255 S298 S303 S445 S515 T140 T403 T506	N504 N74 N90	Signal peptide: M1-G66  Transmembrane domains: V112-V128; I385-L404; L414-I436; Y479-F497  Sugar (and other) transporter domain: A59-F514  Sugar transport proteins signatures: A152-L218  Glucose transporter signature PR00172: V317-Y338; I385-Q405; I416-G439 A449-L467 Y479-L499  Sugar transporter signature PR00171: S68-V78; I168-M187; Y327-F337; I416-L437; G439-F451  SUGAR TRANSPORT PROTEINS DM00135 P22732 132-466: R171-T506  Sugar transporter 1 motif: S371-G386  Sugar transporter 2 motif: I173-R198	SPSCAN  HMMER  HMMER_PFAM  PROFILES SCAN  BLIMPS_PRINTS  BLIMPS_PRINTS  BLAST_DOMO  MOTIFS  MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
26	7611651CD1	905	S105 S140 S145 S200 S26 S283 S288 S435 S55 S617 S653 S671 S698 S721 S735 S811 S819 S826 S844 S876 T13 T170 T202 T220 T301 T326 T363 T377 T433 T469 T625	N218 N457 N589	Transmembrane domains: L300-N318; S394-A412  Transmembrane region cyclic nucleotide gated channel: Y341-I527  Cyclic nucleotide-binding domain: V555-A646  POTASSIUM CHANNEL IONIC CHANNEL PD118772: E649-S902  CAMP RECEPTOR PROTEIN CYCLIC NUCLEOTIDE-BINDING DOMAIN DM01165 I38465 562-948: H413-I418; S420-F685  do POTASSIUM; CHANNEL; KST1; AKT1; DM02383 I38465 353-560: T201-A412  PROTEIN CHANNEL IONIC ION TRANSPORT VOLTAGEGATED P64 CHLORIDE INTRACELLULAR CHLORINE PD017366: Q449-M685	HMMER  HMMER_PFAM  HMMER_PFAM  BLAST_PRODOR  BLAST_DOMO  BLAST_DOMO  BLAST_PRODOR
27	2522075CD1	686	S293 S322 S472 S601 S608 T489 T566 T619 T83	N487		

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
28	7475353CB1	2984	1-605, 2964-2984	7052739LV1 70159545V1 70484059V1 70528817V1 3394211F8 (LJUNGNOT28) 624415R6 (PGANNOT01) 4099055F8 (BRAITUT26) 70483730V1 7249756H2 (PROSTMY01) 5426789F6 (THYMTUT03) 4893528F8 (LIVRTUT12) 1546941R6 (PROSTUT04) 7272275H1 (OVARDIJ01) 4742175F6 (THYMNOR02) GNN.g7160536_000034_00 2 FL180719_00001 GNN.g7523773_000025_00 2 6910236R8 (PITUDIR01) GNN.g7329616_000008_00 2	612 2032 1879 1290 1 2454 456 1265 1682 1262 636 333 1029 1 1 1 169 1 2	1298 2614 2569 1981 603 2984 1097 1882 1846 1807 1171 1053 1694 637 1458 1234 1255 672 957
29	3107278CB1	1846	1-170			
30	7473394CB1	1458	1-1458			
31	7473900CB1	1234				
32	7475045CB1	1255	1-342			
33	7475611CB1	957	1-957			

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
34	7475617CB1	2407	1359-1412, 684-958, 2229-2407	6769264H1 (BRAUNOR01)	1667	2100
				GNN.g7362716_000001_00 2	689	938
				6084383H1 (LUNLUTUT11)	23	664
				5890656F6 (LIVRNON08)	1003	1365
				7695062H1 (LNODTUE01)	1895	2407
				6966295H1 (SKINDIA01)	1340	1895
				97457275	1	492
				60148652D2	799	1149
				5998083F7 (BRAZDIT04)	399	904
				7930723H1 (COLNDIS02)	2135	2745
35	7473314CB1	2767	1-113, 652- 805, 2733- 2767, 2166- 2190	132920F1 (BMARNOT02)	1722	2165
				92207207	2220	2767
				1645009T6 (HEARFET01)	1493	2155
				1710065H1 (PROSNOT16)	2544	2752
				1645009F6 (HEARFET01)	1016	1568
				6767169J1 (BRAUNOR01)	73	774
				6748695H1 (BRAXNOT03)	2409	2750
				3556343H1 (LUNGNOT31)	760	1027
				GNN.g7533975_000016_00 2	1	543
				71753989V1	1678	2182
36	70356714CB1	2182	894-1332	7164046F8 (PLACNOR01)	608	1349
				71759169V1	895	1595
				71757516V1	1484	2171
				5796984F8 (PLACFET04)	1	737

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
37	7611491CB1	2811	826-1021, 2762-2811, 1-176, 1526-1672	8107975J1 (MIXDDIE02) 55030237H1 71749736V1 71749946V1 55030269H1 7088214R8 (BRAUTDR03) 7611491J1 (KIDCTME01) 70211216V1 71152449V1 7708502J1 (PANCNOE02) 71302454V1 7722139J2 (THYRDIE01) 71153625V1 6777664J1 (OVARDIR01) 257274R6 (HNT2RAT01) 257274T6 (HNT2RAT01)	1231 660 2060 2161 520 1 1922 1534 1326 351 1370 1 1420 773 1 715	1970 1315 2624 2811 1232 606 2605 2030 2016 1007 2019 731 2074 1375 571 1340
38	171968CB1	2074	1-773, 1141-2074	768641R6 (LUNGN0T04) 5496021F9 (BRABDIR01) 6355991F8 (LUNGDIS03) 5499076F6 (BRABDIR01) GBI:g7381772_edit1 GBI:g7381772_edit2	1243 5626 5175 5016 1 1378	1581 6027 5816 5432 1377 6027
39	257274CB1	1340	392-1340	7228345H1 (BRAXTDR15) 7664878J1 (UTRSTME01) 4822576H1 (PROSTUT17) 70037119V1 7664878H1 (UTRSTME01) GNN.g6010343_006.edit	1 382 1888 1383 945 1	512 1063 2168 1983 1443 2229
40	6355991CB1	6027	1-558, 846- 1005, 3413- 4373, 1239- 3076			
41	70035348CB1	2168	1-124, 1576-2168, 184-240			
42	7472539CB1	2229	1755-2027, 569-722, 130-409, 1320-1608			

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment (s)	Sequence Fragments	5' Position	3' Position
43	817477CB1	1520	1-151	7765238J1 (URETUE01) FL817477H1_00001 7618286H1 (KIDNTUE01)	921 208 1	1520 1432 573
44	1442166CB1	3950	1-1422, 3633-3950	6765069J1 (BRAUNOR01) 7469439H1 (LUNGNOE02) 71374152V1 7651070J1 (STOMTDE01) 6332268H1 (BRANDIN01) 7176036H1 (BRSTMC01) 7458415H1 (LIVRTUE01) 6836155H1 (BRSTNON02) 1208437R1 (BRSTNOT02) 71374816V1 5884688F8 (LIVERNON08) GBI.g7458720_edit	501 140 1962 751 3413 1333 1379 2107 2746 2681 3252 1	1204 548 2602 1354 3950 1891 2059 2718 3321 3277 3948 232

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment (s)	Sequence Fragments	5' Position	3' Position
45	2311751CB1	5540	1-2744	6762808J1 (BRAUNOR01)	779	1317
				71066032V1	642	1303
				6911060J1 (PITUDIR01)	1219	1795
				5098681F8 (EPIMNON05)	5001	5540
				6766537J1 (BRAUNOR01)	1	747
				4309533H1 (BRAUNOT01)	3898	4261
				7179893H1 (BRAXDIC01)	4839	5348
				6908865J1 (PITUDIR01)	2543	3171
				6769078J1 (BRAUNOR01)	3679	4216
				6770451H1 (BRAUNOR01)	4237	4919
				7467144H1 (LUNGNOE02)	1336	1844
				6765621H1 (BRAUNOR01)	1773	2436
				6763740H1 (BRAUNOR01)	1831	2498
				6893778J1 (BRAITDR03)	3173	3830
46	7472537CB1	2074	1052-1392, 1-462	6889776H1 (BRAITDR03)	4380	4948
				6953905H1 (BRAITDR02)	3099	3798
				6977243H1 (BRAITDR04)	2438	3070
				7984065H1 (UTRSTMC01)	1	541
				g2019266	1752	2074
47	7472546CB1	2259	1350-1638, 596-752, 130-409, 1785-2057	FL7472537_g5815493_g74	428	1852
				06950		
				71400292V1	262	928
				71382167V1	1141	1592
				7218664H1 (COLNTMC01)	1537	1905
				GNN.g6114738_006	1	2259
				4179344F6 (SINITUT03)	261	797
				4669722H1 (SINTNOT24)	2014	2259

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment (s)	Sequence Fragments	5' Position	3' Position
48	7474202CB1	2439	459-822	70218680V2	1909	2439
				70218626V2	1435	2148
				70219176V1	628	1127
				7177066H1 (BRSTMC01)	1596	2259
				70219021V1	542	1003
				7083037H1 (STOMTMR02)	1	619
				70219400V1	1134	1609
49	7476280CB1	2762	1862-1985, 1646-1738, 1-1359, 2113-2184	70219013V1	953	1412
				2756231R6 (THP1AZS08)	1739	2271
				2756231T6 (THP1AZS08)	2242	2736
				g1014431	1386	1648
				g3230934	2349	2762
				GBI.g7622477_edit	1	2762
50	1713377CB1	1897	1-295, 1041-1158	70587572V1	1077	1631
				71875033V1	1459	1897
				1527853T6 (UCMCL5T01)	1343	1854
				71413245V1	1	560
				70587476V1	547	1287
				71413060V1	520	1185



Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
51	5842557CB1	2361	1-688, 2076-2361, 807-885	71052406V1	1311	1916
				70794204V1	377	936
				71412362V1	698	1312
				7695065J1 (LNODTUE01)	1	662
				70730136V1	1820	2361
52	7476643CB1	2032	1-205, 1657-2032	70795377V1	1277	1865
				71207116V1	342	1121
				71197621V1	1355	2032
				71198062V1	1144	1833
				4715941F6 (BRAHCT01)	1	424
53	7611651CB1	2779	2195-2779, 881-936	71205887V1	1093	1666
				71204307V1	429	1135
				71047239V1	1343	1971
				4726692F6 (COLCTUT02)	913	1352
				71047331V1	2011	2614
54	2522075CB1	2430	1-837, 2144-2430	55049229H1	1	817
				71047776V1	2202	2779
				71046696V1	1931	2570
				71048829V1	1331	1678
				55049237J1	212	1118
54				FL2522075_g7717334_g75 92636	1	2061
				7079667H2 (STOMTMR02)	1765	2430

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
28	7475353CB1	PROSNOT14
29	3107278CB1	BRAITUT07
32	7475045CB1	SINITMC01
34	7475617CB1	LIVRNON08
35	7473314CB1	SKINBIT01
36	70356714CB1	PLACNOR01
37	7611491CB1	KIDCTME01
38	171968CB1	BLADNOR01
39	257274CB1	HNT2RAT01
40	6355991CB1	BRABDIR01
41	70035348CB1	LUNGNON03
42	7472539CB1	SINTFEE02
43	817477CB1	KIDNTUE01
44	1442166CB1	BRSTNOT02
45	2311751CB1	BRAUNOR01
46	7472537CB1	PANHTUR01
47	7472546CB1	SINITUT03
48	7474202CB1	BRSTNOT33
49	7476280CB1	THP1AZS08
50	1713377CB1	BMARUNA01
51	5842557CB1	SEMNOT01
52	7476643CB1	LIVRNON08
53	7611651CB1	COLCTUT02
54	2522075CB1	SPLNTUE01

Table 6

Library	Vector	Library Description
BLADNOR01	PCDNA2.1	This random primed library was constructed using RNA isolated from the bladder tissue of an 11-year-old Black male who died from a gunshot wound. Serology was positive for CMV.
BMARUNA01	PSPORT1	Library was constructed using RNA isolated from CD34+ progenitor cells removed from a healthy Black male adult between age 18 and 45, during bilateral bone marrow withdrawal from the posterior iliac crest of the pelvic bone. The CD34+ progenitor cells were isolated from bone marrow mononuclear cells using positive immunomagnetic selection. The patient was a healthy bone marrow donor. The patient was not taking any medications.
BRABDIR01	pINCY	Library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident. Patient history included Huntington's disease, emphysema, and tobacco abuse.
BRAITUT07	pINCY	Library was constructed using RNA isolated from left frontal lobe tumor tissue removed from the brain of a 32-year-old Caucasian male during excision of a cerebral meningeal lesion. Pathology indicated low grade desmoplastic neuronal neoplasm, type not otherwise specified. The lesion formed a firm, circumscribed cyst-associated mass involving white matter and cortex. No definite glial component was evident to suggest a diagnosis of ganglioglioma. Family history included atherosclerotic coronary artery disease.

Table 6 (cont.)

Library	Vector	Library Description
BRAUNOR01	pINCY	This random primed library was constructed using RNA isolated from striatum, globus pallidus and posterior putamen tissue removed from an 81-year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids, bilaterally; microscopic infarcts of the frontal cortex and hippocampus; and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. Grossly, the leptomeninges showed only mild thickening and hyalinization along the superior sagittal sinus. The remainder of the leptomeninges was thin and contained some congested blood vessels. Mild atrophy was found mostly in the frontal poles and lobes, and temporal lobes, bilaterally. Microscopically, there were pairs of Alzheimer type II astrocytes within the deep layers of the neocortex. There was increased satellitosis around neurons in the deep gray matter in the middle frontal cortex. The amygdala contained rare diffuse plaques and neurofibrillary tangles. The posterior hippocampus contained a microscopic area of cystic cavitation with hemosiderin-laden macrophages surrounded by reactive gliosis. Patient history included sepsis, cholangitis, post-operative atelectasis, pneumonia CAD, cardiomegaly due to left ventricular hypertrophy, splenomegaly, arteriolonephrosclerosis, nodular colloidal goiter, emphysema, CHF, hypothyroidism, and peripheral vascular disease.
BRSTNOT02	PSPORT1	Library was constructed using RNA isolated from diseased breast tissue removed from a 55-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated proliferative fibrocystic changes characterized by apocrine metaplasia, sclerosing adenosis, cyst formation, and ductal hyperplasia without atypia. Pathology for the associated tumor tissue indicated an invasive grade 4 mammary adenocarcinoma. Patient history included atrial tachycardia and a benign neoplasm. Family history included cardiovascular and cerebrovascular disease.
BRSTNOT33	pINCY	Library was constructed using RNA isolated from right breast tissue removed from a 46-year-old Caucasian female during unilateral extended simple mastectomy with breast reconstruction. Pathology for the associated tumor tissue indicated invasive grade 3 adenocarcinoma, ductal type, with apocrine features, nuclear grade 3 forming a mass in the outer quadrant. There was greater than 50% intraductal component. Patient history included breast cancer.

Table 6 (cont.)

Library	Vector	Library Description
COLCTUT02	PINCY	Library was constructed using RNA isolated from colon tumor tissue removed from the cecum of a 30-year-old Caucasian female during partial colectomy, open liver biopsy, incidental appendectomy, and permanent colostomy. Pathology indicated carcinoid tumor (grade 1 neuroendocrine carcinoma) arising in the terminal ileum, forming a mass in the right colon. Patient history included chronic sinus infections and endometriosis. Family history included hyperlipidemia, anxiety, upper lobe lung cancer, stomach cancer, liver cancer, and cirrhosis.
HNT2RAT01	PBLUESCRIPT	Library was constructed at Stratagene (STR937231), using RNA isolated from the hNT2 cell line (derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor). Cells were treated with retinoic acid for 24 hours
KIDCTME01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from kidney cortex tissue removed from a 65-year-old male during nephroureterectomy. Pathology indicated the margins of resection were free of involvement. Pathology for the matched tumor tissue indicated grade 3 renal cell carcinoma, clear cell type, forming a variegated multicystic mass situated within the mid-portion of the kidney. The tumor invaded deeply into but not through the renal capsule.
KIDNTUE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from kidney tumor tissue removed from a 46-year-old Caucasian male during nephroureterectomy. Pathology indicated grade 2 renal cell carcinoma, clear-cell type, forming a mass in the upper pole. The patient presented with kidney cancer, backache, headache, malignant hypertension, nausea, and vomiting. Previous surgeries included repair of indirect inguinal hernia. Patient medications included Lasix, Inderal, and Procardia. Family history included cerebrovascular accident in the mother; acute myocardial infarction and atherosclerotic coronary artery disease in the father; and type II diabetes in the sibling(s).

Table 6 (cont.)

Library	Vector	Library Description
LIVRNON08	pINCY	This normalized library was constructed from 5.7 million independent clones from a pooled liver tissue library. Starting RNA was made from pooled liver tissue removed from a 4-year-old Hispanic male who died from anoxia and a 16 week female fetus who died after 16-weeks gestation from anencephaly. Serologies were positive for cytomegalovirus in the 4-year-old. Patient history included asthma in the 4-year-old. Family history included taking daily prenatal vitamins and mitral valve prolapse in the mother of the fetus. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
LIVRNON08	pINCY	This normalized library was constructed from 5.7 million independent clones from a pooled liver tissue library. Starting RNA was made from pooled liver tissue removed from a 4-year-old Hispanic male who died from anoxia and a 16 week female fetus who died after 16-weeks gestation from anencephaly. Serologies were positive for cytomegalovirus in the 4-year-old. Patient history included asthma in the 4-year-old. Family history included taking daily prenatal vitamins and mitral valve prolapse in the mother of the fetus. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
LUNGNON03	PSPORT1	This normalized library was constructed from 2.56 million independent clones from a lung tissue library. RNA was made from lung tissue removed from the left lobe a 58-year-old Caucasian male during a segmental lung resection. Pathology for the associated tumor tissue indicated a metastatic grade 3 (of 4) osteosarcoma. Patient history included soft tissue cancer, secondary cancer of the lung, prostate cancer, and an acute duodenal ulcer with hemorrhage. Patient also received radiation therapy to the retroperitoneum. Family history included prostate cancer, breast cancer, and acute leukemia. The normalization and hybridization conditions were adapted from Soares et al., PNAS (1994) 91:9228; Swaroop et al., NAR (1991) 19:1954; and Bonaldo et al., Genome Research (1996) 6:791.

Table 6 (cont.)

Library	Vector	Library Description
PANHTUR01	PBK-CMV	This random primed library was constructed using RNA isolated from pancreatic tumor tissue removed from a 65-year-old female. Pathology indicated well-differentiated neuroendocrine carcinoma (islet cell tumor), nuclear grade 1, forming a dominant mass in the distal pancreas. Multiple smaller tumor nodules were immediately adjacent to the main mass. The liver showed metastatic grade 1 islet cell tumor, forming multiple nodules. Multiple (4) pericholedochal lymph nodes contained metastatic grade 1 islet cell tumor.
PLACNOR01	PCDNA2.1	This random primed library was constructed using pooled cDNA from two different donors. cDNA was generated using mRNA isolated from placental tissue removed from a Caucasian fetus (donor A), who died after 16 weeks' gestation from fetal demise and hydrocephalus and from placental tissue removed from a Caucasian male fetus (donor B), who died after 18 weeks' gestation from fetal demise. Patient history for donor A included umbilical cord wrapped around the head (3 times) and the shoulders (1 time). Serology was positive for anti-CMV and remaining serologies were negative. Family history included multiple pregnancies and live births, and an abortion in the mother. Serology was negative for donor B.
PROSNOT14	PINCY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 60-year-old Caucasian male during radical prostatectomy and regional lymph node excision. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+4). The patient presented with elevated prostate specific antigen (PSA). Patient history included a kidney cyst and hematuria. Family history included benign hypertension, cerebrovascular disease, and arteriosclerotic coronary artery disease.
SEMVNOT01	PINCY	Library was constructed using RNA isolated from seminal vesicle tissue removed from a 58-year-old Caucasian male during radical prostatectomy. Pathology for the associated tumor tissue indicated adenocarcinoma (Gleason grade 3+2) of the prostate. Adenofibromatous hyperplasia was also present. The patient presented with elevated prostate specific antigen (PSA). Family history included a malignant breast neoplasm.

Table 6 (cont.)

Library	Vector	Library Description
SINITMC01	pINCY	<p>This large size-fractionated library was constructed using pooled cDNA from two donors. cDNA was generated using mRNA isolated from ileum tissue removed from a 30-year-old Caucasian female (donor A) during partial colectomy, open liver biopsy, and permanent colostomy, and from ileum tissue removed from a 70-year-old Caucasian female (donor B) during right hemicolectomy, open liver biopsy, sigmoidoscopy, colonoscopy, and permanent colostomy. Pathology for the matched tumor tissue (donor A) indicated carcinoma tumor (grade 1 neuroendocrine carcinoma) arising in the terminal ileum. The tumor permeated through the ileal wall into the mesenteric fat and extended into the adherent cecum, where tumor extended through the bowel wall up to the mucosal surface. Multiple lymph nodes were positive for tumor. Additional (2) lymph nodes were also involved by direct tumor extension. Pathology for donor B indicated a non-tumorous margin of ileum. Pathology for the matched tumor (donor B) indicated invasive grade 2 adenocarcinoma forming an ulcerated mass, situated distal to the ileocecal valve. The tumor invaded through the muscularis propria just into the serosal adipose tissue. One regional lymph node was positive for a microfocus of metastatic adenocarcinoma. Donor A presented with flushing and unspecified abdominal/pelvic symptoms. Patient history included endometriosis, and tobacco and alcohol abuse. Donor B's history included a malignant breast neoplasm, type II diabetes, hyperlipidemia, viral hepatitis, an unspecified thyroid disorder, osteoarthritis, and a malignant skin neoplasm. Donor B's medication included tamoxifen.</p>
SINITUT03	pINCY	<p>Library was constructed using RNA isolated from ileal tumor tissue obtained from a 49-year-old Caucasian female during destruction of peritoneal tissue, peritoneal adhesiolysis, ileum resection, and permanent colostomy. Pathology indicated grade 4 adenocarcinoma. Patient history included benign hypertension. Previous surgeries included total abdominal hysterectomy, bilateral salpingo-oophorectomy, regional lymph node excision, an incidental appendectomy, and dilation and curettage. Family history included benign hypertension, cerebrovascular disease, hyperlipidemia, atherosclerotic coronary artery disease, hyperlipidemia, type II diabetes, and stomach cancer.</p>



Table 6 (cont.)

Library	Vector	Library Description
SINTFEE02	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from small intestine tissue removed from a Caucasian male fetus who died from Patau's syndrome (trisomy 13) at 20-weeks' gestation. Serology was negative.
SKINBIT01	PINCY	Library was constructed using RNA isolated from diseased skin tissue of the left lower leg. Patient history included erythema nodosum of the left lower leg.
SPLNTUE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from spleen tumor tissue removed from a 28-year-old male during total splenectomy. Pathology indicated malignant lymphoma, diffuse large cell type, B-cell phenotype with abundant reactive T-cells and marked granulomatous response involving the spleen, where it formed approximately 45 nodules, liver, and multiple lymph nodes.
THP1AZS08	PSPORT1	This subtracted THP-1 promonocyte cell line library was constructed using 5.76 x 1e6 clones from a 5-aza-2'-deoxycytidine (AZ) treated THP-1 cell library. Starting RNA was made from THP-1 promonocyte cells treated for three days with 0.8 micromolar AZ. The donor had acute monocytic leukemia. The hybridization probe for subtraction was derived from a similarly constructed library, made from 1 microgram of polyA RNA isolated from untreated THP-1 cells. 5.76 million clones from the AZ-treated THP-1 cell library were then subjected to two rounds of subtractive hybridization with 5 million clones from the untreated THP-1 cell library. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR (1991) 19:1954, and Bonaldo et al., Genome Research (1996) 6:791.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.0E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value= 1.0E-3 or less Signal peptide hits: Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score > GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
  - a) a polypeptide comprising an amino acid sequence selected from the group consisting of  
5 SEQ ID NO:1-27,
  - b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical  
to an amino acid sequence selected from the group consisting of SEQ ID NO:1-27,
  - c) a biologically active fragment of a polypeptide having an amino acid sequence selected  
from the group consisting of SEQ ID NO:1-27, and
  - 10 d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from  
the group consisting of SEQ ID NO:1-27.
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-27.
- 15 3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID  
20 NO:28-54.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a  
polynucleotide of claim 3.
- 25 7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method for producing a polypeptide of claim 1, the method comprising:
  - 30 a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell  
is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a  
promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
  - b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide selected from the group consisting of:

- a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54,
- b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:28-54,
- c) a polynucleotide complementary to a polynucleotide of a),
- d) a polynucleotide complementary to a polynucleotide of b), and
- e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
- b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
- b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-27.

18. A method for treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional TRICH, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- 10 c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- 20 b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

- 25 a) treating a biological sample containing nucleic acids with the test compound;
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim
- 30 11 or fragment thereof;
- c) quantifying the amount of hybridization complex; and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

29. A diagnostic test for a condition or disease associated with the expression of TRICH in a biological sample comprising the steps of:

- a) combining the biological sample with an antibody of claim 10, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex; and
- b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

30. The antibody of claim 10, wherein the antibody is:

- a) a chimeric antibody,
- b) a single chain antibody,
- c) a Fab fragment,
- d) a F(ab')<sub>2</sub> fragment, or
- e) a humanized antibody.

31. A composition comprising an antibody of claim 10 and an acceptable excipient.

32. A method of diagnosing a condition or disease associated with the expression of TRICH in a subject, comprising administering to said subject an effective amount of the composition of claim

31.

33. A composition of claim 31, wherein the antibody is labeled.

34. A method of diagnosing a condition or disease associated with the expression of TRICH in a subject, comprising administering to said subject an effective amount of the composition of claim 33.

35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 10 comprising:

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, or an immunogenic fragment thereof, under conditions to elicit an antibody response;
- b) isolating antibodies from said animal; and



c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27.

5 36. An antibody produced by a method of claim 35.

37. A composition comprising the antibody of claim 36 and a suitable carrier.

38. A method of making a monoclonal antibody with the specificity of the antibody of claim  
10 10 comprising:

a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27, or an immunogenic fragment thereof, under conditions to elicit an antibody response;

b) isolating antibody producing cells from the animal;

15 c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells;

d) culturing the hybridoma cells; and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27.

20 39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

25 41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.

30 43. A method for detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27 in a sample, comprising the steps of:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding

of the antibody and the polypeptide; and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27 in the sample.

5 44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27 from a sample, the method comprising:

a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-27.

45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

15 47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

20 49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

50. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

51. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

25 52. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

53. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

30 54. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

55. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.
57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.
58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.
59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.
60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.
61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.
62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.
63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.
64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.
65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.
66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:22.
67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:23.
68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:24.
69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:25.
70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:26.
71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:27.
72. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:28.

73. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:29.
74. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:30.
- 5 75. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:31.
76. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:32.
77. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:33.
- 10 78. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:34.
79. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:35.
- 15 80. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:36.
81. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:37.
82. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:38.
- 20 83. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:39.
84. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:40.
- 25 85. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:41.
86. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:42.
87. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:43.
- 30 88. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:44.
89. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:45.

90. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:46.
91. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:47.
- 5 92. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:48.
93. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:49.
94. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:50.
- 10 95. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:51.
96. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:52.
- 15 97. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:53.
98. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:54.

<110> INCYTE GENOMICS, INC.  
 THORNTON, Michael  
 WALIA, Narinder K.  
 YUE, Henry  
 NGUYEN, Dannel B.  
 LAL, Preeti  
 GANDHI, Ameena R.  
 TRIBOULEY, Catherine M.  
 YAO, Monique G.  
 RAMKUMAR, Jayalaxmi  
 AU-YOUNG, Janice  
 LU, Yan  
 TANG, Y. Tom  
 AZIMZAI, Yalda  
 BRUNS, Christopher M.  
 GRIFFIN, Jennifer A.  
 YANG, Junming  
 BAUGHN, Mariah R.  
 SANJANWALA, Madhu S.  
 RAUMANN, Brigitte E.  
 LEE, Ernestine A.  
 HAFALIA, April  
 GREENE, Barrie D.  
 KHAN, Farrah A.  
 KEARNEY, Liam  
 ELLIOTT, Vicky S.  
 SEILHAMER, Jeffrey J.  
 POLICKY, Jennifer L.  
 BOROWSKY, Mark L.  
 BURFORD, Neil  
 DING, Li  
 LU, Dyung Aina M.  
 HILLMAN, Jennifer L.

<120> TRANSPORTERS AND ION CHANNELS

<130> PI-0103 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/208,424; 60/209,001; 60/210,588; 60/212,335; 60/213,747; 60/215,391  
 <151> 2000-05-26; 2000-06-01; 2000-06-08; 2000-06-16; 2000-06-22; 2000-06-29

<160> 54

<170> PERL Program

<210> 1

<211> 842

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7475353CD1

<400> 1

Met	Val	Thr	Val	Gly	Asn	Tyr	Cys	Glu	Ala	Glu	Gly	Pro	Val	Gly
1				5				10						15
Pro	Ala	Trp	Met	Gln	Asp	Gly	Leu	Ser	Pro	Cys	Phe	Phe	Phe	Thr
				20				25						30
Leu	Val	Pro	Ser	Thr	Arg	Met	Ala	Leu	Gly	Thr	Leu	Ala	Leu	Val
				35				40						45

Leu	Ala	Leu	Pro	Cys	Arg	Arg	Arg	Glu	Arg	Pro	Ala	Gly	Ala	Asp
				50					55					60
Ser	Leu	Ser	Trp	Gly	Ala	Gly	Pro	Arg	Ile	Ser	Pro	Tyr	Val	Leu
				65					70					75
Gln	Leu	Leu	Leu	Ala	Thr	Leu	Gln	Ala	Ala	Leu	Pro	Leu	Ala	Gly
				80					85					90
Leu	Ala	Gly	Arg	Val	Gly	Thr	Ala	Arg	Gly	Ala	Pro	Leu	Pro	Ser
				95					100					105
Tyr	Leu	Leu	Leu	Ala	Ser	Val	Leu	Glu	Ser	Leu	Ala	Gly	Ala	Cys
				110					115					120
Gly	Leu	Trp	Leu	Leu	Val	Val	Glu	Arg	Ser	Gln	Ala	Arg	Gln	Arg
				125					130					135
Leu	Ala	Met	Gly	Ile	Trp	Ile	Lys	Phe	Arg	His	Ser	Pro	Gly	Leu
				140					145					150
Leu	Leu	Leu	Trp	Thr	Val	Ala	Phe	Ala	Ala	Glu	Asn	Leu	Ala	Leu
				155					160					165
Val	Ser	Trp	Asn	Ser	Pro	Gln	Trp	Trp	Trp	Ala	Arg	Ala	Asp	Leu
				170					175					180
Gly	Gln	Gln	Val	Gln	Phe	Ser	Leu	Trp	Val	Leu	Arg	Tyr	Val	Val
				185					190					195
Ser	Gly	Gly	Leu	Phe	Val	Leu	Gly	Leu	Trp	Ala	Pro	Gly	Leu	Arg
				200					205					210
Pro	Gln	Ser	Tyr	Thr	Leu	Gln	Val	His	Glu	Glu	Asp	Gln	Asp	Val
				215					220					225
Glu	Arg	Ser	Gln	Val	Arg	Ser	Ala	Ala	Gln	Gln	Ser	Thr	Trp	Arg
				230					235					240
Asp	Phe	Gly	Arg	Lys	Leu	Arg	Leu	Leu	Ser	Gly	Tyr	Leu	Trp	Pro
				245					250					255
Arg	Gly	Ser	Pro	Ala	Leu	Gln	Leu	Val	Val	Leu	Ile	Cys	Leu	Gly
				260					265					270
Leu	Met	Gly	Leu	Glu	Arg	Ala	Leu	Asn	Val	Leu	Val	Pro	Ile	Phe
				275					280					285
Tyr	Arg	Asn	Ile	Val	Asn	Leu	Leu	Thr	Glu	Lys	Ala	Pro	Trp	Asn
				290					295					300
Ser	Leu	Ala	Trp	Thr	Val	Thr	Ser	Tyr	Val	Phe	Leu	Lys	Phe	Leu
				305					310					315
Gln	Gly	Gly	Gly	Thr	Gly	Ser	Thr	Gly	Phe	Val	Ser	Asn	Leu	Arg
				320					325					330
Thr	Phe	Leu	Trp	Ile	Arg	Val	Gln	Gln	Phe	Thr	Ser	Arg	Arg	Val
				335					340					345
Glu	Leu	Leu	Ile	Phe	Ser	His	Leu	His	Glu	Leu	Ser	Leu	Arg	Trp
				350					355					360
His	Leu	Gly	Arg	Arg	Thr	Gly	Glu	Val	Leu	Arg	Ile	Ala	Asp	Arg
				365					370					375
Gly	Thr	Ser	Ser	Val	Thr	Gly	Leu	Leu	Ser	Tyr	Leu	Val	Phe	Asn
				380					385					390
Val	Ile	Pro	Thr	Leu	Ala	Asp	Ile	Ile	Ile	Gly	Ile	Ile	Tyr	Phe
				395					400					405
Ser	Met	Phe	Phe	Asn	Ala	Trp	Phe	Gly	Leu	Ile	Val	Phe	Leu	Cys
				410					415					420
Met	Ser	Leu	Tyr	Leu	Thr	Leu	Thr	Ile	Val	Val	Thr	Glu	Trp	Arg
				425					430					435
Thr	Lys	Phe	Arg	Arg	Ala	Met	Asn	Thr	Gln	Glu	Asn	Ala	Thr	Arg
				440					445					450
Ala	Arg	Ala	Val	Asp	Ser	Leu	Leu	Asn	Phe	Glu	Thr	Val	Lys	Tyr
				455					460					465
Tyr	Asn	Ala	Glu	Ser	Tyr	Glu	Val	Glu	Arg	Tyr	Arg	Glu	Ala	Ile
				470					475					480
Ile	Lys	Tyr	Gln	Gly	Leu	Glu	Trp	Lys	Ser	Ser	Ala	Ser	Leu	Val
				485					490					495
Leu	Leu	Asn	Gln	Thr	Gln	Asn	Leu	Val	Ile	Gly	Leu	Gly	Leu	Leu
				500					505					510
Ala	Gly	Ser	Leu	Leu	Cys	Ala	Tyr	Phe	Val	Thr	Glu	Gln	Lys	Leu

Gln Val Gly Asp	515	Tyr Val Leu Phe Gly	520	Thr Tyr Ile Ile Gln Leu	525
Tyr Met Pro Leu	530	Asn Trp Phe Gly Thr	535	Tyr Tyr Arg Met Ile Gln	540
Thr Asn Phe Ile	545	Asp Met Glu Asn Met	550	Phe Asp Leu Leu Lys Glu	555
Glu Thr Glu Val	560	Lys Asp Leu Pro Gly	565	Ala Gly Pro Leu Arg Phe	570
Gln Lys Gly Arg	575	Ile Glu Phe Glu Asn	580	Val His Phe Ser Tyr Ala	585
Asp Gly Arg Glu	590	Thr Leu Gln Asp Val	595	Ser Phe Thr Val Met Pro	600
Gly Gln Thr Leu	605	Ala Leu Val Gly Pro	610	Ser Gly Ala Gly Lys Ser	615
Thr Ile Leu Arg	620	Leu Leu Phe Arg Phe	625	Tyr Asp Ile Ser Ser Gly	630
Cys Ile Arg Ile	635	Asp Gly Gln Asp Ile	640	Ser Gln Val Thr Gln Ala	645
Ser Leu Arg Ser	650	His Ile Gly Val Val	655	Pro Gln Asp Thr Val Leu	660
Phe Asn Asp Thr	665	Ile Ala Asp Asn Ile	670	Arg Tyr Gly Arg Val Thr	675
Ala Gly Asn Asp	680	Glu Val Glu Ala Ala	685	Ala Gln Ala Ala Gly Ile	690
His Asp Ala Ile	695	Met Ala Phe Pro Glu	700	Gly Tyr Arg Thr Gln Val	705
Gly Glu Arg Gly	710	Leu Lys Leu Ser Gly	715	Gly Glu Lys Gln Arg Val	720
Ala Ile Ala Arg	725	Thr Ile Leu Lys Ala	730	Pro Gly Ile Ile Leu Leu	735
Asp Glu Ala Thr	740	Ser Ala Leu Asp Thr	745	Ser Asn Glu Arg Ala Ile	750
Gln Ala Ser Leu	755	Ala Lys Val Cys Ala	760	Asn Arg Thr Thr Ile Val	765
Val Ala His Arg	770	Leu Ser Thr Val Val	775	Asn Ala Asp Gln Ile Leu	780
Val Ile Lys Asp	785	Gly Cys Ile Val Glu	790	Arg Gly Arg His Glu Ala	795
Leu Leu Ser Arg	800	Gly Gly Val Tyr Ala	805	Asp Met Trp Gln Leu Gln	810
Gln Gly Gln Glu	815	Glu Thr Ser Glu Asp	820	Thr Lys Pro Gln Thr Met	825
Glu Arg	830		835		840

<210> 2  
 <211> 461  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 3107278CD1

<400> 2  
 Met Pro Gly Arg Ser Ile Ser Leu Ser Ser Pro Tyr Trp Trp Ile  
 1 5 10 15  
 Asn Leu Trp Tyr Leu Ile Thr Gly Cys Ile Ala Asp Trp Val Gly  
 20 25 30  
 Arg Arg Pro Val Leu Leu Phe Ser Ile Ile Phe Ile Leu Ile Phe  
 35 40 45  
 Gly Leu Thr Val Ala Leu Ser Val Asn Val Thr Met Phe Ser Thr



Leu	Arg	Phe	Phe	Glu	Gly	Phe	Cys	Leu	Ala	Gly	Ile	Ile	Leu	Thr	50	55	60
				65					70						75		
Leu	Tyr	Ala	Leu	Arg	Ile	Glu	Leu	Cys	Pro	Pro	Gly	Lys	Arg	Phe	80	85	90
Met	Ile	Thr	Met	Val	Ala	Ser	Phe	Val	Ala	Met	Ala	Gly	Gln	Phe	95	100	105
Leu	Met	Pro	Gly	Leu	Ala	Ala	Leu	Cys	Arg	Asp	Trp	Gln	Val	Leu	110	115	120
Gln	Ala	Leu	Ile	Ile	Cys	Pro	Phe	Leu	Leu	Met	Leu	Leu	Tyr	Trp	125	130	135
Ser	Ile	Phe	Pro	Glu	Ser	Leu	Arg	Trp	Leu	Met	Ala	Thr	Gln	Gln	140	145	150
Phe	Glu	Ser	Ala	Lys	Arg	Leu	Ile	Leu	His	Phe	Thr	Gln	Lys	Asn	155	160	165
Arg	Met	Asn	Pro	Glu	Gly	Asp	Ile	Lys	Gly	Val	Ile	Pro	Glu	Leu	170	175	180
Glu	Lys	Glu	Leu	Ser	Arg	Arg	Pro	Lys	Lys	Val	Cys	Ile	Val	Lys	185	190	195
Val	Val	Gly	Thr	Arg	Asn	Leu	Trp	Lys	Asn	Ile	Val	Val	Leu	Cys	200	205	210
Val	Asn	Ser	Leu	Thr	Gly	Tyr	Gly	Ile	His	His	Cys	Phe	Ala	Arg	215	220	225
Ser	Met	Met	Gly	His	Glu	Val	Lys	Val	Pro	Leu	Leu	Glu	Asn	Phe	230	235	240
Tyr	Ala	Asp	Tyr	Tyr	Thr	Thr	Ala	Ser	Ile	Ala	Leu	Val	Ser	Cys	245	250	255
Leu	Ala	Met	Cys	Val	Val	Val	Arg	Phe	Leu	Gly	Arg	Arg	Gly	Gly	260	265	270
Leu	Leu	Leu	Phe	Met	Ile	Leu	Thr	Ala	Leu	Ala	Ser	Leu	Leu	Gln	275	280	285
Leu	Gly	Leu	Leu	Asn	Leu	Ile	Gly	Lys	Tyr	Ser	Gln	His	Pro	Asp	290	295	300
Ser	Gly	Met	Ser	Asp	Ser	Val	Lys	Asp	Lys	Phe	Ser	Ile	Ala	Phe	305	310	315
Ser	Ile	Val	Gly	Met	Phe	Ala	Ser	His	Ala	Val	Gly	Ser	Leu	Ser	320	325	330
Val	Phe	Phe	Cys	Ala	Glu	Ile	Thr	Pro	Thr	Val	Ile	Arg	Cys	Gly	335	340	345
Gly	Leu	Gly	Leu	Val	Leu	Ala	Ser	Ala	Gly	Phe	Gly	Met	Leu	Thr	350	355	360
Ala	Pro	Ile	Ile	Glu	Leu	His	Asn	Gln	Lys	Gly	Tyr	Phe	Leu	His	365	370	375
His	Ile	Ile	Phe	Ala	Cys	Cys	Thr	Leu	Ile	Cys	Ile	Ile	Cys	Ile	380	385	390
Leu	Leu	Leu	Pro	Glu	Ser	Arg	Asp	Gln	Asn	Leu	Pro	Glu	Asn	Ile	395	400	405
Ser	Asn	Gly	Glu	His	Tyr	Thr	Arg	Gln	Pro	Leu	Leu	Pro	His	Lys	410	415	420
Lys	Gly	Glu	Gln	Pro	Leu	Leu	Leu	Thr	Asn	Ala	Glu	Leu	Lys	Asp	425	430	435
Tyr	Ser	Gly	Leu	His	Asp	Ala	Ala	Ala	Ala	Gly	Asp	Thr	Leu	Pro	440	445	450
Glu	Gly	Ala	Thr	Ala	Asn	Gly	Met	Lys	Ala	Met					455	460	

<210> 3  
 <211> 485  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature

&lt;223&gt; Incyte ID No: 7473394CD1

&lt;400&gt; 3

Met	Gln	Asn	Ile	Thr	Lys	Glu	Phe	Gly	Thr	Phe	Lys	Ala	Asn	Asp
1				5					10					15
Asn	Ile	Asn	Leu	Gln	Val	Lys	Ala	Gly	Glu	Ile	His	Ala	Leu	Leu
				20					25					30
Gly	Glu	Asn	Gly	Ala	Gly	Lys	Ser	Thr	Leu	Met	Asn	Val	Leu	Ser
				35					40					45
Gly	Leu	Leu	Glu	Pro	Thr	Ser	Gly	Lys	Ile	Leu	Met	Arg	Gly	Lys
				50					55					60
Glu	Val	Gln	Ile	Thr	Ser	Pro	Thr	Lys	Ala	Asn	Gln	Leu	Gly	Ile
				65					70					75
Gly	Met	Val	His	Gln	His	Phe	Met	Leu	Val	Asp	Ala	Phe	Thr	Val
				80					85					90
Thr	Glu	Asn	Ile	Val	Leu	Gly	Ser	Glu	Pro	Ser	Arg	Ala	Gly	Met
				95					100					105
Leu	Asp	His	Lys	Lys	Ala	Arg	Lys	Glu	Ile	Gln	Lys	Val	Ser	Glu
				110					115					120
Gln	Tyr	Gly	Leu	Ser	Val	Asn	Pro	Asp	Ala	Tyr	Val	Arg	Asp	Ile
				125					130					135
Ser	Val	Gly	Met	Glu	Gln	Arg	Val	Glu	Ile	Leu	Lys	Thr	Leu	Tyr
				140					145					150
Arg	Gly	Ala	Asp	Val	Leu	Ile	Phe	Asp	Glu	Pro	Thr	Ala	Val	Leu
				155					160					165
Thr	Pro	Gln	Glu	Ile	Asp	Glu	Leu	Ile	Val	Ile	Met	Lys	Glu	Leu
				170					175					180
Val	Lys	Glu	Gly	Lys	Ser	Ile	Ile	Leu	Ile	Thr	His	Lys	Leu	Asp
				185					190					195
Glu	Ile	Lys	Ala	Val	Ala	Asp	Arg	Cys	Thr	Val	Ile	Arg	Arg	Gly
				200					205					210
Lys	Gly	Ile	Gly	Thr	Val	Asn	Val	Lys	Asp	Val	Thr	Ser	Gln	Gln
				215					220					225
Leu	Ala	Asp	Met	Met	Val	Gly	Arg	Ala	Val	Ser	Phe	Lys	Thr	Met
				230					235					240
Lys	Lys	Glu	Ala	Lys	Pro	Gln	Glu	Val	Val	Leu	Ser	Ile	Glu	Asn
				245					250					255
Leu	Val	Val	Lys	Glu	Asn	Arg	Gly	Leu	Glu	Ala	Val	Lys	Asn	Leu
				260					265					270
Asn	Leu	Glu	Val	Arg	Ala	Gly	Glu	Val	Leu	Gly	Ile	Ala	Gly	Ile
				275					280					285
Asp	Gly	Asn	Gly	Gln	Ser	Glu	Leu	Ile	Gln	Ala	Leu	Thr	Gly	Leu
				290					295					300
Arg	Lys	Ala	Glu	Ser	Gly	His	Ile	Lys	Leu	Lys	Gly	Glu	Asp	Ile
				305					310					315
Thr	Asn	Lys	Lys	Pro	Arg	Lys	Ile	Thr	Glu	His	Gly	Val	Gly	His
				320					325					330
Val	Pro	Glu	Asp	Arg	His	Lys	Tyr	Gly	Leu	Val	Leu	Asp	Met	Thr
				335					340					345
Leu	Ser	Glu	Asn	Ile	Ala	Leu	Gln	Thr	Tyr	His	Gln	Lys	Pro	Tyr
				350					355					360
Ser	Lys	Asn	Gly	Met	Leu	Asn	Tyr	Ser	Val	Ile	Asn	Glu	His	Ala
				365					370					375
Arg	Glu	Leu	Ile	Glu	Glu	Tyr	Asp	Val	Arg	Thr	Thr	Asn	Glu	Leu
				380					385					390
Val	Pro	Ala	Lys	Ala	Leu	Ser	Gly	Gly	Asn	Gln	Gln	Lys	Ala	Ile
				395					400					405
Ile	Ala	Arg	Ile	Val	Asp	Arg	Asp	Pro	Asp	Leu	Leu	Ile	Val	Ala
				410					415					420
Asn	Pro	Thr	Arg	Gly	Leu	Asp	Val	Gly	Ala	Ile	Glu	Phe	Ile	His
				425					430					435
Lys	Arg	Leu	Ile	Glu	Gln	Arg	Asp	Lys	Tyr	Lys	Ala	Val	Leu	Leu
				440					445					450

Ile Ser Phe Glu Leu Glu Glu Ile Leu Asn Val Ser Asp Arg Ile  
 455 460 465  
 Ala Val Ile His Glu Gly Glu Ile Val Gly Ile Val Asp Pro Lys  
 470 475 480  
 Glu Thr Ser Glu Asn  
 485

<210> 4  
 <211> 301  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7473900CD1

<400> 4  
 Met Lys Ser Gly Pro Gly Ile Gln Ala Ala Ile Asp Leu Thr Ala  
 1 5 10 15  
 Gly Ala Ala Gly Gly Thr Ala Cys Val Leu Thr Gly Gln Pro Phe  
 20 25 30  
 Asp Thr Ile Lys Val Lys Met Gln Thr Phe Pro Asp Leu Tyr Lys  
 35 40 45  
 Gly Leu Thr Asp Cys Phe Leu Lys Thr Tyr Ala Gln Val Gly Leu  
 50 55 60  
 Arg Gly Phe Tyr Lys Gly Thr Gly Pro Ala Leu Met Ala Tyr Val  
 65 70 75  
 Ala Glu Asn Ser Val Leu Phe Met Cys Tyr Gly Phe Cys Gln Gln  
 80 85 90  
 Phe Val Arg Lys Val Ala Gly Met Asp Lys Gln Ala Lys Leu Ser  
 95 100 105  
 Asp Leu Gln Thr Ala Ala Ala Gly Ser Phe Ala Ser Ala Phe Ala  
 110 115 120  
 Ala Leu Ala Leu Cys Pro Thr Glu Leu Val Lys Cys Arg Leu Gln  
 125 130 135  
 Thr Met Tyr Glu Met Glu Met Ser Gly Lys Ile Ala Lys Ser His  
 140 145 150  
 Asn Thr Ile Trp Ser Val Val Lys Gly Ile Leu Lys Lys Asp Gly  
 155 160 165  
 Pro Leu Gly Phe Tyr His Gly Leu Ser Ser Thr Leu Leu Gln Glu  
 170 175 180  
 Val Pro Gly Tyr Phe Phe Phe Phe Gly Gly Tyr Glu Leu Ser Arg  
 185 190 195  
 Ser Phe Phe Ala Ser Gly Arg Ser Lys Asp Glu Leu Gly Pro Val  
 200 205 210  
 His Leu Met Leu Ser Gly Gly Val Ala Gly Ile Cys Leu Trp Leu  
 215 220 225  
 Val Val Phe Pro Val Asp Cys Ile Lys Ser Arg Ile Gln Val Leu  
 230 235 240  
 Ser Met Tyr Gly Lys Gln Ala Gly Phe Ile Gly Thr Leu Leu Ser  
 245 250 255  
 Val Val Arg Asn Glu Gly Ile Val Ala Leu Tyr Ser Gly Leu Lys  
 260 265 270  
 Ala Thr Met Ile Arg Ala Ile Pro Ala Asn Gly Ala Leu Phe Val  
 275 280 285  
 Ala Tyr Glu Tyr Ser Arg Lys Met Met Met Lys Gln Leu Glu Ala  
 290 295 300  
 Tyr

<210> 5  
 <211> 304  
 <212> PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7475045CD1

&lt;400&gt; 5

Met	Glu	Thr	Val	Pro	Pro	Ala	Val	Asp	Leu	Val	Leu	Gly	Ala	Ser
1				5					10					15
Ala	Cys	Cys	Leu	Ala	Cys	Val	Phe	Thr	Asn	Pro	Leu	Glu	Val	Val
				20					25					30
Lys	Thr	Arg	Leu	Gln	Leu	Gln	Gly	Glu	Leu	Gln	Ala	Arg	Gly	Thr
				35					40					45
Tyr	Pro	Arg	Pro	Tyr	His	Gly	Phe	Ile	Ala	Ser	Val	Ala	Ala	Val
				50					55					60
Ala	Arg	Ala	Asp	Gly	Leu	Trp	Gly	Leu	Gln	Lys	Gly	Leu	Ala	Ala
				65					70					75
Gly	Leu	Leu	Tyr	Gln	Gly	Leu	Met	Asn	Gly	Val	Arg	Phe	Tyr	Cys
				80					85					90
Tyr	Ser	Leu	Ala	Cys	Gln	Ala	Gly	Leu	Thr	Gln	Gln	Pro	Gly	Gly
				95					100					105
Thr	Val	Val	Ala	Gly	Ala	Val	Ala	Gly	Ala	Leu	Gly	Ala	Phe	Val
				110					115					120
Gly	Ser	Pro	Ala	Tyr	Leu	Ile	Lys	Thr	Gln	Leu	Gln	Ala	Gln	Thr
				125					130					135
Val	Ala	Ala	Val	Ala	Val	Gly	His	Gln	His	Asn	His	Gln	Thr	Val
				140					145					150
Leu	Gly	Ala	Leu	Glu	Thr	Ile	Trp	Arg	Gln	Gln	Gly	Leu	Leu	Gly
				155					160					165
Leu	Trp	Gln	Gly	Val	Gly	Gly	Ala	Val	Pro	Arg	Val	Met	Val	Gly
				170					175					180
Ser	Ala	Ala	Gln	Leu	Ala	Thr	Phe	Ala	Ser	Ala	Lys	Ala	Trp	Val
				185					190					195
Gln	Lys	Gln	Gln	Trp	Leu	Pro	Glu	Asp	Ser	Trp	Leu	Val	Ala	Leu
				200					205					210
Ala	Gly	Gly	Met	Ile	Ser	Ser	Ile	Ala	Val	Val	Val	Val	Met	Thr
				215					220					225
Pro	Phe	Asp	Val	Val	Ser	Thr	Arg	Leu	Tyr	Asn	Gln	Pro	Val	Asp
				230					235					240
Thr	Ala	Gly	Arg	Gly	Gln	Leu	Tyr	Gly	Gly	Leu	Thr	Asp	Cys	Met
				245					250					255
Val	Lys	Ile	Trp	Arg	Gln	Glu	Gly	Pro	Leu	Ala	Leu	Tyr	Lys	Gly
				260					265					270
Leu	Gly	Pro	Ala	Tyr	Leu	Arg	Leu	Gly	Pro	His	Thr	Ile	Leu	Ser
				275					280					285
Met	Leu	Phe	Trp	Asp	Glu	Leu	Arg	Lys	Leu	Ala	Gly	Arg	Ala	Gln
				290					295					300
His	Lys	Gly	Thr											

&lt;210&gt; 6

&lt;211&gt; 278

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7475611CD1

&lt;400&gt; 6

Met	Ser	Ala	Lys	Val	Leu	Leu	Ser	Thr	Glu	His	Leu	Tyr	Ala	Thr
1				5					10					15
His	Pro	Gly	Arg	Pro	Met	Val	Leu	Thr	Asp	Val	Asn	Val	Ser	Phe

Arg	Ala	Gly	Val	Arg	Val	Ala	Ile	Leu	Gly	Ala	Asn	Gly	Ser	Gly	20	25	30
Lys	Thr	Thr	Leu	Met	Arg	Cys	Leu	Ser	Gly	Ser	Leu	Lys	Pro	Ala	35	40	45
Lys	Gly	His	Val	Lys	Arg	Gly	Asp	Ile	Val	Val	Ser	Tyr	Gly	Arg	50	55	60
Ala	Gln	Leu	Arg	Glu	His	Arg	Arg	Ala	Val	Gln	Leu	Val	Leu	Gln	65	70	75
Asp	Pro	Asp	Asp	Gln	Leu	Phe	Ser	Ala	Asp	Val	Ser	Gln	Asp	Val	80	85	90
Ser	Phe	Gly	Pro	Met	Asn	Met	Gly	Leu	Lys	Val	Asp	Glu	Val	Arg	95	100	105
Asp	Arg	Val	Ser	Glu	Ser	Leu	Glu	Leu	Leu	Gly	Ala	Ser	His	Leu	110	115	120
Ala	Glu	Arg	Ala	Thr	Tyr	Gln	Leu	Ser	Tyr	Gly	Glu	Arg	Lys	Arg	125	130	135
Val	Ala	Val	Ala	Gly	Ala	Val	Ala	Met	Arg	Pro	Asp	Leu	Leu	Leu	140	145	150
Leu	Asp	Glu	Pro	Thr	Ala	Gly	Leu	Asp	Pro	Val	Gly	Val	Thr	Gln	155	160	165
Met	Leu	Glu	Ala	Leu	Asp	Arg	Leu	Arg	Asp	His	Gly	Thr	Thr	Val	170	175	180
Ala	Met	Ala	Thr	His	Asp	Val	Asp	Leu	Ala	Leu	Ala	Trp	Ala	Gln	185	190	195
Glu	Ala	Leu	Val	Val	Val	Asp	Gly	Gln	Val	His	Gln	Gly	Pro	Ile	200	205	210
Gly	Glu	Leu	Leu	Ala	Asp	Ala	Asp	Thr	Val	Gly	Arg	Ala	His	Leu	215	220	225
His	Leu	Pro	Trp	Pro	Leu	Glu	Leu	Ala	Arg	Arg	Leu	Gly	Val	Arg	230	235	240
Asp	Leu	Pro	Arg	Thr	Met	Asp	Asp	Val	Val	Ala	Met	Leu	Ser	Asp	245	250	255
Asn	Pro	Ser	Pro	Ala	Pro	Ser	Asn								260	265	270
															275		

&lt;210&gt; 7

&lt;211&gt; 673

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7475617CD1

&lt;400&gt; 7

Met	Glu	Glu	Leu	Ala	Thr	Glu	Lys	Glu	Ala	Glu	Glu	Ser	His	Arg	1	5	10	15
Gln	Asp	Ser	Val	Ser	Leu	Leu	Thr	Phe	Ile	Leu	Leu	Leu	Thr	Leu	20	25	30	35
Thr	Ile	Leu	Thr	Ile	Trp	Leu	Phe	Lys	His	Arg	Arg	Val	Arg	Phe	35	40	45	50
Leu	His	Glu	Thr	Gly	Leu	Ala	Met	Ile	Tyr	Gly	Leu	Ile	Val	Gly	50	55	60	65
Val	Ile	Leu	Arg	Tyr	Gly	Thr	Pro	Ala	Thr	Ser	Gly	Arg	Asp	Lys	65	70	75	80
Ser	Leu	Ser	Cys	Thr	Gln	Glu	Asp	Arg	Ala	Phe	Ser	Thr	Leu	Leu	80	85	90	95
Val	Asn	Val	Ser	Gly	Lys	Phe	Phe	Glu	Tyr	Thr	Leu	Lys	Gly	Glu	95	100	105	110
Ile	Ser	Pro	Gly	Lys	Ile	Asn	Ser	Val	Glu	Gln	Asn	Asp	Met	Leu	110	115	120	125
Arg	Lys	Val	Thr	Phe	Asp	Pro	Glu	Val	Phe	Phe	Asn	Ile	Leu	Leu	125	130	135	140

	125		130		135
Pro Pro Ile Ile	Phe His Ala Gly Tyr	Ser Leu Lys Lys Arg	His		
	140		145		150
Phe Phe Arg Asn	Leu Gly Ser Ile Leu	Ala Tyr Ala Phe Leu	Gly		
	155		160		165
Thr Ala Val Ser	Cys Phe Ile Ile Gly	Asn Leu Met Tyr Gly	Val		
	170		175		180
Val Lys Leu Met	Lys Ile Met Gly Gln	Leu Ser Asp Lys Phe	Tyr		
	185		190		195
Tyr Thr Asp Cys	Leu Phe Phe Gly Ala	Ile Ile Ser Ala Thr	Asp		
	200		205		210
Pro Val Thr Val	Leu Ala Ile Phe Asn	Glu Leu His Ala Asp	Val		
	215		220		225
Asp Leu Tyr Ala	Leu Leu Phe Gly Glu	Ser Val Leu Asn Asp	Ala		
	230		235		240
Val Ala Ile Val	Leu Ser Ser Ser Ile	Val Ala Tyr Gln Pro	Ala		
	245		250		255
Gly Leu Asn Thr	His Ala Phe Asp Ala	Ala Ala Phe Phe Lys	Ser		
	260		265		270
Val Gly Ile Phe	Leu Gly Ile Phe Ser	Gly Ser Phe Thr Met	Gly		
	275		280		285
Ala Val Thr Gly	Val Val Thr Ala Leu	Val Thr Lys Phe Thr	Lys		
	290		295		300
Leu His Cys Phe	Pro Leu Leu Glu Thr	Ala Leu Phe Phe Leu	Met		
	305		310		315
Ser Trp Ser Thr	Phe Leu Leu Ala Glu	Ala Cys Gly Phe Thr	Gly		
	320		325		330
Val Val Ala Val	Leu Phe Cys Gly Ile	Thr Gln Ala His Tyr	Thr		
	335		340		345
Tyr Asn Asn Leu	Ser Val Glu Ser Arg	Ser Arg Thr Lys Gln	Leu		
	350		355		360
Phe Glu Val Leu	His Phe Leu Ala Glu	Asn Phe Ile Phe Ser	Tyr		
	365		370		375
Met Gly Leu Ala	Leu Phe Thr Phe Gln	Lys His Val Phe Ser	Pro		
	380		385		390
Ile Phe Ile Ile	Gly Ala Phe Val Ala	Ile Phe Leu Gly Arg	Ala		
	395		400		405
Ala His Ile Tyr	Pro Leu Ser Phe Phe	Leu Asn Leu Gly Arg	Arg		
	410		415		420
His Lys Ile Gly	Trp Asn Phe Gln His	Met Met Met Phe Ser	Gly		
	425		430		435
Leu Arg Gly Ala	Met Ala Phe Ala Leu	Ala Ile Arg Asp Thr	Ala		
	440		445		450
Ser Tyr Ala Arg	Gln Met Met Phe Thr	Thr Thr Leu Leu Ile	Val		
	455		460		465
Phe Phe Thr Val	Trp Ile Ile Gly Gly	Gly Thr Thr Pro Met	Leu		
	470		475		480
Ser Trp Leu Asn	Ile Arg Val Gly Val	Glu Glu Pro Ser Glu	Glu		
	485		490		495
Asp Gln Asn Glu	His His Trp Gln Tyr	Phe Arg Val Gly Val	Asp		
	500		505		510
Pro Asp Gln Asp	Pro Pro Pro Asn Asn	Asp Ser Phe Gln Val	Leu		
	515		520		525
Gln Gly Asp Gly	Pro Asp Ser Ala Arg	Gly Asn Arg Thr Lys	Gln		
	530		535		540
Glu Ser Ala Trp	Ile Phe Arg Leu Trp	Tyr Ser Phe Asp His	Asn		
	545		550		555
Tyr Leu Lys Pro	Ile Leu Thr His Ser	Gly Pro Pro Leu Thr	Thr		
	560		565		570
Thr Leu Pro Ala	Trp Cys Gly Leu Leu	Ala Arg Cys Leu Thr	Ser		
	575		580		585
Pro Gln Val Tyr	Asp Asn Gln Glu Pro	Leu Arg Glu Glu Asp	Ser		
	590		595		600

Asp	Phe	Ile	Leu	Thr	Glu	Gly	Asp	Leu	Thr	Leu	Thr	Tyr	Gly	Asp
				605					610					615
Ser	Thr	Val	Thr	Ala	Asn	Gly	Ser	Ser	Ser	Ser	His	Thr	Ala	Ser
				620					625					630
Thr	Ser	Leu	Glu	Gly	Ser	Arg	Arg	Thr	Lys	Ser	Ser	Ser	Glu	Glu
				635					640					645
Val	Leu	Glu	Arg	Asp	Leu	Gly	Met	Gly	Asp	Gln	Lys	Val	Ser	Ser
				650					655					660
Arg	Gly	Thr	Arg	Leu	Val	Phe	Pro	Leu	Glu	Asp	Asn	Ala		
				665					670					

&lt;210&gt; 8

&lt;211&gt; 576

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7473314CD1

&lt;400&gt; 8

Met	Glu	Gly	Ser	Gly	Gly	Gly	Ala	Gly	Glu	Arg	Ala	Pro	Leu	Leu
1				5					10					15
Gly	Ala	Arg	Arg	Ala	Ala	Ala	Ala	Ala	Ala	Gly	Ala	Phe	Ala	
				20					25					30
Gly	Arg	Arg	Ala	Ala	Cys	Gly	Ala	Val	Leu	Leu	Thr	Glu	Leu	Leu
				35					40					45
Glu	Arg	Ala	Ala	Phe	Tyr	Gly	Ile	Thr	Ser	Asn	Leu	Val	Leu	Phe
				50					55					60
Leu	Asn	Gly	Ala	Pro	Phe	Cys	Trp	Glu	Gly	Ala	Gln	Ala	Ser	Glu
				65					70					75
Ala	Leu	Leu	Leu	Phe	Met	Gly	Leu	Thr	Tyr	Leu	Gly	Ser	Pro	Phe
				80					85					90
Gly	Gly	Trp	Leu	Ala	Asp	Ala	Arg	Leu	Gly	Arg	Ala	Arg	Ala	Ile
				95					100					105
Leu	Leu	Ser	Leu	Ala	Leu	Tyr	Leu	Leu	Gly	Met	Leu	Ala	Phe	Pro
				110					115					120
Leu	Leu	Ala	Ala	Pro	Ala	Thr	Arg	Ala	Ala	Leu	Cys	Gly	Ser	Ala
				125					130					135
Arg	Leu	Leu	Asn	Cys	Thr	Ala	Pro	Gly	Pro	Asp	Ala	Ala	Ala	Arg
				140					145					150
Cys	Cys	Ser	Pro	Ala	Thr	Phe	Ala	Gly	Leu	Val	Leu	Val	Gly	Leu
				155					160					165
Gly	Val	Ala	Thr	Val	Lys	Ala	Asn	Ile	Thr	Pro	Phe	Gly	Ala	Asp
				170					175					180
Gln	Val	Lys	Asp	Arg	Gly	Pro	Glu	Ala	Thr	Arg	Arg	Phe	Phe	Asn
				185					190					195
Trp	Phe	Tyr	Trp	Ser	Ile	Asn	Leu	Gly	Ala	Ile	Leu	Ser	Leu	Gly
				200					205					210
Gly	Ile	Ala	Tyr	Ile	Gln	Gln	Asn	Val	Ser	Phe	Val	Thr	Gly	Tyr
				215					220					225
Ala	Ile	Pro	Thr	Val	Cys	Val	Gly	Leu	Ala	Phe	Val	Ala	Phe	Leu
				230					235					240
Cys	Gly	Gln	Ser	Val	Phe	Ile	Thr	Lys	Pro	Pro	Asp	Gly	Ser	Ala
				245					250					255
Phe	Thr	Asp	Met	Phe	Lys	Ile	Leu	Thr	Tyr	Ser	Cys	Cys	Ser	Gln
				260					265					270
Lys	Arg	Ser	Gly	Glu	Arg	Gln	Ser	Asn	Gly	Glu	Gly	Ile	Gly	Val
				275					280					285
Phe	Gln	Gln	Ser	Ser	Lys	Gln	Ser	Leu	Phe	Asp	Ser	Cys	Lys	Met
				290					295					300
Ser	His	Gly	Gly	Pro	Phe	Thr	Glu	Glu	Lys	Val	Glu	Asp	Val	Lys
				305					310					315

<400>	9														
Met	Ala	Phe	Ser	Lys	Leu	Leu	Glu	Gln	Ala	Gly	Gly	Val	Gly	Leu	
1				5					10					15	
Phe	Gln	Thr	Leu	Gln	Val	Leu	Thr	Phe	Ile	Leu	Pro	Cys	Leu	Met	
				20					25					30	
Ile	Pro	Ser	Gln	Met	Leu	Leu	Glu	Asn	Phe	Ser	Ala	Ala	Ile	Pro	
				35					40					45	
Gly	His	Arg	Cys	Trp	Thr	His	Met	Leu	Asp	Asn	Gly	Ser	Ala	Val	
				50					55					60	
Ser	Thr	Asn	Met	Thr	Pro	Lys	Ala	Leu	Leu	Thr	Ile	Ser	Ile	Pro	
				65					70					75	
Pro	Gly	Pro	Asn	Gln	Gly	Pro	His	Gln	Cys	Arg	Arg	Phe	Arg	Gln	
				80					85					90	
Pro	Gln	Trp	Gln	Leu	Leu	Asp	Pro	Asn	Ala	Thr	Ala	Thr	Ser	Trp	
				95					100					105	
Ser	Glu	Ala	Asp	Thr	Glu	Pro	Cys	Val	Asp	Gly	Trp	Val	Tyr	Asp	
				110					115					120	



Arg	Ser	Val	Phe	Thr	Ser	Thr	Ile	Val	Ala	Lys	Trp	Asp	Leu	Val
				125					130					135
Cys	Ser	Ser	Gln	Gly	Leu	Lys	Pro	Leu	Ser	Gln	Ser	Ile	Phe	Met
				140					145					150
Ser	Gly	Ile	Leu	Val	Gly	Ser	Phe	Ile	Trp	Gly	Leu	Leu	Ser	Tyr
				155					160					165
Arg	Phe	Gly	Arg	Lys	Pro	Met	Leu	Ser	Trp	Cys	Cys	Leu	Gln	Leu
				170					175					180
Ala	Val	Ala	Gly	Thr	Ser	Thr	Ile	Phe	Ala	Pro	Thr	Phe	Val	Ile
				185					190					195
Tyr	Cys	Gly	Leu	Arg	Phe	Val	Ala	Ala	Phe	Gly	Met	Ala	Gly	Ile
				200					205					210
Phe	Leu	Ser	Ser	Leu	Thr	Leu	Met	Val	Glu	Trp	Thr	Thr	Thr	Ser
				215					220					225
Arg	Arg	Ala	Val	Thr	Met	Thr	Val	Val	Gly	Cys	Ala	Phe	Ser	Ala
				230					235					240
Gly	Gln	Ala	Ala	Leu	Gly	Gly	Leu	Ala	Phe	Ala	Leu	Arg	Asp	Trp
				245					250					255
Arg	Thr	Leu	Gln	Leu	Ala	Ala	Ser	Val	Pro	Phe	Phe	Ala	Ile	Ser
				260					265					270
Leu	Ile	Ser	Trp	Trp	Leu	Pro	Glu	Ser	Ala	Arg	Trp	Leu	Ile	Ile
				275					280					285
Lys	Gly	Lys	Pro	Asp	Gln	Ala	Leu	Gln	Glu	Leu	Arg	Lys	Val	Ala
				290					295					300
Arg	Ile	Asn	Gly	His	Lys	Glu	Ala	Lys	Asn	Leu	Thr	Ile	Glu	Val
				305					310					315
Leu	Met	Ser	Ser	Val	Lys	Glu	Glu	Val	Ala	Ser	Ala	Lys	Glu	Pro
				320					325					330
Arg	Ser	Val	Leu	Asp	Leu	Phe	Cys	Val	Pro	Val	Leu	Arg	Trp	Arg
				335					340					345
Ser	Cys	Ala	Met	Leu	Val	Val	Asn	Phe	Ser	Leu	Leu	Ile	Ser	Tyr
				350					355					360
Tyr	Gly	Leu	Val	Phe	Asp	Leu	Gln	Ser	Leu	Gly	Arg	Asp	Ile	Phe
				365					370					375
Leu	Leu	Gln	Ala	Leu	Phe	Gly	Ala	Val	Asp	Phe	Leu	Gly	Arg	Ala
				380					385					390
Thr	Thr	Ala	Leu	Leu	Leu	Ser	Phe	Leu	Gly	Arg	Arg	Thr	Ile	Gln
				395					400					405
Ala	Gly	Ser	Gln	Ala	Met	Ala	Gly	Leu	Ala	Ile	Leu	Ala	Asn	Met
				410					415					420
Leu	Val	Pro	Gln	Asp	Leu	Gln	Thr	Leu	Arg	Val	Val	Phe	Ala	Val
				425					430					435
Leu	Gly	Lys	Gly	Cys	Phe	Gly	Ile	Ser	Leu	Thr	Cys	Leu	Thr	Ile
				440					445					450
Tyr	Lys	Ala	Glu	Leu	Phe	Pro	Thr	Pro	Val	Arg	Met	Thr	Ala	Asp
				455					460					465
Gly	Ile	Leu	His	Thr	Val	Gly	Arg	Leu	Gly	Ala	Met	Met	Gly	Pro
				470					475					480
Leu	Ile	Leu	Met	Ser	Arg	Gln	Ala	Leu	Pro	Leu	Leu	Pro	Pro	Leu
				485					490					495
Leu	Tyr	Gly	Val	Ile	Ser	Ile	Ala	Ser	Ser	Leu	Val	Val	Leu	Phe
				500					505					510
Phe	Leu	Pro	Glu	Thr	Gln	Gly	Leu	Pro	Leu	Pro	Asp	Thr	Ile	Gln
				515					520					525
Asp	Leu	Glu	Ser	Gln	Lys	Ser	Thr	Ala	Ala	Gln	Gly	Asn	Arg	Gln
				530					535					540
Glu	Ala	Val	Thr	Val	Glu	Ser	Thr	Ser	Leu					
				545					550					

&lt;210&gt; 10

&lt;211&gt; 559

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7611491CD1

&lt;400&gt; 10

Met	Arg	Arg	Gln	Asp	Ser	Arg	Gly	Asn	Thr	Val	Leu	His	Ala	Leu	1	5	10	15
Val	Ala	Ile	Ala	Asp	Asn	Thr	Arg	Glu	Asn	Thr	Lys	Phe	Val	Thr	20	25	30	
Lys	Met	Tyr	Asp	Leu	Leu	Leu	Leu	Lys	Cys	Ala	Arg	Leu	Phe	Pro	35	40	45	
Asp	Ser	Asn	Leu	Glu	Ala	Val	Leu	Asn	Asn	Asp	Gly	Leu	Ser	Pro	50	55	60	
Leu	Met	Met	Met	Ala	Ala	Lys	Thr	Gly	Lys	Ile	Gly	Ile	Phe	Gln	65	70	75	
His	Ile	Ile	Arg	Arg	Glu	Val	Thr	Asp	Glu	Asp	Thr	Arg	His	Leu	80	85	90	
Ser	Arg	Lys	Phe	Lys	Asp	Trp	Ala	Tyr	Gly	Pro	Val	Tyr	Ser	Ser	95	100	105	
Leu	Tyr	Asp	Leu	Ser	Ser	Leu	Asp	Thr	Cys	Gly	Glu	Glu	Ala	Ser	110	115	120	
Val	Leu	Glu	Ile	Leu	Val	Tyr	Asn	Ser	Lys	Ile	Glu	Asn	Arg	His	125	130	135	
Glu	Met	Leu	Ala	Val	Glu	Pro	Ile	Asn	Glu	Leu	Leu	Arg	Asp	Lys	140	145	150	
Trp	Arg	Lys	Phe	Gly	Ala	Val	Ser	Phe	Tyr	Ile	Asn	Val	Val	Ser	155	160	165	
Tyr	Leu	Cys	Ala	Met	Val	Ile	Phe	Thr	Leu	Thr	Ala	Tyr	Tyr	Gln	170	175	180	
Pro	Leu	Glu	Gly	Thr	Pro	Pro	Tyr	Pro	Tyr	Arg	Thr	Thr	Val	Asp	185	190	195	
Tyr	Leu	Arg	Leu	Ala	Gly	Glu	Val	Ile	Thr	Leu	Phe	Thr	Gly	Val	200	205	210	
Leu	Phe	Phe	Phe	Thr	Asn	Ile	Lys	Asp	Leu	Phe	Met	Lys	Lys	Cys	215	220	225	
Pro	Gly	Val	Asn	Ser	Leu	Phe	Ile	Asp	Gly	Ser	Phe	Gln	Leu	Leu	230	235	240	
Tyr	Phe	Ile	Tyr	Ser	Val	Leu	Val	Ile	Val	Ser	Ala	Ala	Leu	Tyr	245	250	255	
Leu	Ala	Gly	Ile	Glu	Ala	Tyr	Leu	Ala	Val	Met	Val	Phe	Ala	Leu	260	265	270	
Val	Leu	Gly	Trp	Met	Asn	Ala	Leu	Tyr	Phe	Thr	Arg	Gly	Leu	Lys	275	280	285	
Leu	Thr	Gly	Thr	Tyr	Ser	Ile	Met	Ile	Gln	Lys	Ile	Leu	Phe	Lys	290	295	300	
Asp	Leu	Phe	Arg	Phe	Leu	Leu	Val	Tyr	Leu	Leu	Phe	Met	Ile	Gly	305	310	315	
Tyr	Ala	Ser	Ala	Leu	Val	Ser	Leu	Leu	Asn	Pro	Cys	Ala	Asn	Met	320	325	330	
Lys	Val	Cys	Asn	Glu	Asp	Gln	Thr	Asn	Cys	Thr	Val	Pro	Thr	Tyr	335	340	345	
Pro	Ser	Cys	Arg	Asp	Ser	Glu	Thr	Phe	Ser	Thr	Phe	Leu	Leu	Asp	350	355	360	
Leu	Phe	Lys	Leu	Thr	Ile	Gly	Met	Gly	Asp	Leu	Glu	Met	Leu	Ser	365	370	375	
Ser	Thr	Lys	Tyr	Pro	Val	Val	Phe	Ile	Ile	Leu	Leu	Val	Thr	Tyr	380	385	390	
Ile	Ile	Leu	Thr	Phe	Val	Leu	Leu	Leu	Asn	Met	Leu	Ile	Ala	Leu	395	400	405	
Met	Gly	Glu	Thr	Val	Gly	Gln	Val	Ser	Lys	Glu	Ser	Lys	His	Ile	410	415	420	
Trp	Lys	Leu	Gln	Trp	Ala	Thr	Thr	Ile	Leu	Asp	Ile	Glu	Arg	Ser	425	430	435	

Phe	Pro	Val	Phe	Leu	Arg	Lys	Ala	Phe	Arg	Ser	Gly	Glu	Met	Val	
				440					445					450	
Thr	Val	Gly	Lys	Ser	Ser	Asp	Gly	Thr	Pro	Asp	Arg	Arg	Trp	Cys	
				455					460					465	
Phe	Arg	Val	Asp	Glu	Val	Asn	Trp	Ser	His	Trp	Asn	Gln	Asn	Leu	
				470					475					480	
Gly	Ile	Ile	Asn	Glu	Asp	Pro	Gly	Lys	Asn	Glu	Thr	Tyr	Gln	Tyr	
				485					490					495	
Tyr	Gly	Phe	Ser	His	Thr	Val	Gly	Arg	Leu	Arg	Arg	Asp	Arg	Trp	
				500					505					510	
Ser	Ser	Val	Val	Pro	Arg	Val	Val	Glu	Leu	Asn	Lys	Asn	Ser	Asn	
				515					520					525	
Pro	Asp	Glu	Val	Val	Val	Pro	Leu	Asp	Ser	Thr	Gly	Asn	Pro	Arg	
				530					535					540	
Cys	Asp	Gly	His	Gln	Gln	Gly	Tyr	Pro	Arg	Lys	Trp	Arg	Thr	Asp	
				545					550					555	
Asp	Ala	Pro	Leu												

<210> 11  
 <211> 181  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 171968CD1

<400>	11														
Met	Phe	His	His	Gln	Gln	Ala	Tyr	Cys	Leu	Ala	Pro	Phe	Asp	Leu	
1				5					10					15	
Ile	Lys	Val	Arg	Leu	Gln	Asn	Gln	Thr	Glu	Pro	Arg	Ala	Gln	Pro	
				20					25					30	
Gly	Ser	Pro	Pro	Pro	Arg	Tyr	Gln	Gly	Pro	Val	His	Cys	Ala	Ala	
				35					40					45	
Ser	Ile	Phe	Arg	Glu	Glu	Gly	Pro	Arg	Gly	Leu	Phe	Arg	Gly	Ala	
				50					55					60	
Trp	Ala	Leu	Thr	Leu	Arg	Asp	Thr	Pro	Thr	Val	Gly	Ile	Tyr	Phe	
				65					70					75	
Ile	Thr	Tyr	Glu	Gly	Leu	Cys	Arg	Gln	Tyr	Thr	Pro	Glu	Gly	Gln	
				80					85					90	
Asn	Pro	Ser	Ser	Ala	Thr	Val	Leu	Val	Ala	Gly	Gly	Phe	Ala	Gly	
				95					100					105	
Ile	Ala	Ser	Trp	Val	Ala	Ala	Thr	Pro	Leu	Asp	Val	Ile	Lys	Ser	
				110					115					120	
Arg	Met	Gln	Met	Asp	Gly	Leu	Arg	Arg	Arg	Val	Tyr	Gln	Gly	Met	
				125					130					135	
Leu	Asp	Cys	Met	Val	Ser	Ser	Ile	Arg	Gln	Glu	Gly	Leu	Gly	Val	
				140					145					150	
Phe	Phe	Arg	Gly	Val	Thr	Ile	Asn	Ser	Ala	Arg	Ala	Phe	Pro	Val	
				155					160					165	
Asn	Ala	Val	Thr	Phe	Leu	Ser	Tyr	Glu	Tyr	Leu	Leu	Arg	Trp	Trp	
				170					175					180	
Gly															

<210> 12  
 <211> 124  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature

&lt;223&gt; Incyte ID No: 257274CD1

&lt;400&gt; 12

Met	Cys	Ser	Gly	Leu	Leu	Glu	Leu	Leu	Leu	Pro	Ile	Trp	Leu	Ser
1				5					10					15
Trp	Thr	Leu	Gly	Thr	Arg	Gly	Ser	Glu	Pro	Arg	Ser	Val	Asn	Asp
				20					25					30
Pro	Gly	Asn	Met	Ser	Phe	Val	Lys	Glu	Thr	Val	Asp	Lys	Leu	Leu
				35					40					45
Lys	Gly	Tyr	Asp	Ile	Arg	Leu	Arg	Pro	Asp	Phe	Gly	Gly	Pro	Pro
				50					55					60
Val	Cys	Val	Gly	Met	Asn	Ile	Asp	Ile	Ala	Ser	Ile	Asp	Met	Val
				65					70					75
Ser	Glu	Val	Asn	Met	Arg	Phe	Trp	Leu	Gln	Glu	Arg	Gly	Thr	Lys
				80					85					90
Thr	Val	Val	Cys	Ala	Phe	Gln	Gly	Cys	Leu	Cys	Gly	Phe	Ser	Lys
				95					100					105
Ala	Ala	Ser	Trp	Thr	Gly	Arg	Pro	Gly	Pro	Gly	Thr	Ala	Ser	Leu
				110					115					120
Cys	Pro	Arg	Cys											

&lt;210&gt; 13

&lt;211&gt; 2009

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6355991CD1

&lt;400&gt; 13

Met	Glu	Gln	Thr	Val	Leu	Val	Pro	Pro	Gly	Pro	Asp	Ser	Phe	Asn
1				5					10					15
Phe	Phe	Thr	Arg	Glu	Ser	Leu	Ala	Ala	Ile	Glu	Arg	Arg	Ile	Ala
				20					25					30
Glu	Glu	Lys	Ala	Lys	Asn	Pro	Lys	Pro	Asp	Lys	Lys	Asp	Asp	Asp
				35					40					45
Glu	Asn	Gly	Pro	Lys	Pro	Asn	Ser	Asp	Leu	Glu	Ala	Gly	Lys	Asn
				50					55					60
Leu	Pro	Phe	Ile	Tyr	Gly	Asp	Ile	Pro	Pro	Glu	Met	Val	Ser	Glu
				65					70					75
Pro	Leu	Glu	Asp	Leu	Asp	Pro	Tyr	Tyr	Ile	Asn	Lys	Gln	Thr	Phe
				80					85					90
Ile	Val	Leu	Asn	Lys	Gly	Lys	Ala	Ile	Phe	Arg	Phe	Ser	Ala	Thr
				95					100					105
Ser	Ala	Leu	Tyr	Ile	Leu	Thr	Pro	Phe	Asn	Pro	Leu	Arg	Lys	Ile
				110					115					120
Ala	Ile	Lys	Ile	Leu	Val	His	Ser	Leu	Phe	Ser	Met	Leu	Ile	Met
				125					130					135
Cys	Thr	Ile	Leu	Thr	Asn	Cys	Val	Phe	Met	Thr	Met	Ser	Asn	Pro
				140					145					150
Pro	Asp	Trp	Thr	Lys	Asn	Val	Glu	Tyr	Thr	Phe	Thr	Gly	Ile	Tyr
				155					160					165
Thr	Phe	Glu	Ser	Leu	Ile	Lys	Ile	Ile	Ala	Arg	Gly	Phe	Cys	Leu
				170					175					180
Glu	Asp	Phe	Thr	Phe	Leu	Arg	Asp	Pro	Trp	Asn	Trp	Leu	Asp	Phe
				185					190					195
Thr	Val	Ile	Thr	Phe	Ala	Tyr	Val	Thr	Glu	Phe	Val	Asp	Leu	Gly
				200					205					210
Asn	Val	Ser	Ala	Leu	Arg	Thr	Phe	Arg	Val	Leu	Arg	Ala	Leu	Lys
				215					220					225
Thr	Ile	Ser	Val	Ile	Pro	Gly	Leu	Lys	Thr	Ile	Val	Gly	Ala	Leu

Ile Gln Ser Val	230	Lys Lys Leu Ser Asp	235	Val Met Ile Leu Thr	240
	245		250		255
Phe Cys Leu Ser	260	Val Phe Ala Leu Ile	265	Gly Leu Gln Leu Phe	270
	275	Asn Lys Cys Ile Gln	280	Pro Pro Thr Asn	285
Gly Asn Leu Arg	290	His Ser Ile Glu Lys	295	Asn Ile Thr Val Asn	300
	305	Ile Asn Glu Thr Val	310	Phe Glu Phe Asp Trp	315
Asn Gly Thr Leu	320	Asp Ser Gly Tyr His	325	Tyr Phe Leu Glu Gly	330
Ser Tyr Ile Gln	335	Leu Cys Gly Asn Ser	340	Ser Asp Ala Gly Gln	345
	350	Met Cys Val Lys Ala	355	Gly Arg Asn Pro Asn	360
Leu Asp Ala Leu	365	Phe Asp Thr Phe Ser	370	Trp Ala Phe Leu Ser	375
	380	Thr Gln Asp Phe Trp	385	Glu Asn Leu Tyr Gln	390
Phe Arg Leu Met	395	Ala Gly Lys Thr Tyr	400	Met Ile Phe Phe Val	405
	410	Gly Ser Phe Tyr Leu	415	Ile Asn Leu Ile Leu	420
Val Ile Phe Leu	425	Ala Tyr Glu Glu Gln	430	Asn Gln Ala Thr Leu	435
	440	Lys Glu Ala Glu Phe	445	Gln Gln Met Ile Glu	450
Glu Ala Glu Gln	455	Gln Glu Ala Ala Gln	460	Gln Ala Ala Thr Ala	465
	470	Ala Ser Arg Glu Pro Ser	475	Ala Ala Gly Arg Leu	480
Leu Lys Lys Gln	485	Glu Ala Ser Lys Leu	490	Ser Ser Lys Ser Ala	495
	500	Arg Arg Lys Lys Arg	505	Lys Gln Lys Glu Gln	510
Ala Ser Glu His	515	Lys Asp Glu Asp Glu	520	Phe Gln Lys Ser Glu	525
	530	Arg Arg Lys Gly Phe	535	Arg Phe Ser Ile Glu	540
Glu Asp Ser Ile	545	Tyr Glu Lys Arg Tyr	550	Ser Ser Pro His Gln	555
	560	Arg Gly Ser Leu Phe	565	Ser Pro Arg Arg Asn	570
Asn Arg Leu Thr	575	Phe Ser Phe Arg Gly	580	Arg Ala Lys Asp Val	585
	590	Phe Ala Asp Asp Glu	595	His Ser Thr Phe Glu	600
Ser Glu Asn Asp	605	Arg Asp Ser Leu Phe	610	Val Pro Arg Arg His	615
	620	Ser Asn Leu Ser Gln	625	Thr Ser Arg Ser Ser	630
Asn Glu Ser Arg	635	Phe Pro Ala Asn Gly	640	Lys Met His Ser Thr	645
	650	Val Val Ser Leu Val	655	Gly Gly Pro Ser Val	660
Glu Arg Arg Asn	665	Gln Leu Leu Pro	670	Glu Val Ile Ile Asp	675
	680	Asp Asn Gly Thr Thr	685	Thr Glu Thr Glu Met	690
Met Leu Ala Val	695	Ser Ser Phe His Val	700	Ser Met Asp Phe Leu	705

Asp	Pro	Ser	Gln	Arg	Gln	Arg	Ala	Met	Ser	Ile	Ala	Ser	Ile	Leu
				710					715					720
Thr	Asn	Thr	Val	Glu	Glu	Leu	Glu	Glu	Ser	Arg	Gln	Lys	Cys	Pro
				725					730					735
Pro	Cys	Trp	Tyr	Lys	Phe	Ser	Asn	Ile	Phe	Leu	Ile	Trp	Asp	Cys
				740					745					750
Ser	Pro	Tyr	Trp	Leu	Lys	Val	Lys	His	Val	Val	Asn	Leu	Val	Val
				755					760					765
Met	Asp	Pro	Phe	Val	Asp	Leu	Ala	Ile	Thr	Ile	Cys	Ile	Val	Leu
				770					775					780
Asn	Thr	Leu	Phe	Met	Ala	Met	Glu	His	Tyr	Pro	Met	Thr	Asp	His
				785					790					795
Phe	Asn	Asn	Val	Leu	Thr	Val	Gly	Asn	Leu	Val	Phe	Thr	Gly	Ile
				800					805					810
Phe	Thr	Ala	Glu	Met	Phe	Leu	Lys	Ile	Ile	Ala	Met	Asp	Pro	Tyr
				815					820					825
Tyr	Tyr	Phe	Gln	Glu	Gly	Trp	Asn	Ile	Phe	Asp	Gly	Phe	Ile	Val
				830					835					840
Thr	Leu	Ser	Leu	Val	Glu	Leu	Gly	Leu	Ala	Asn	Val	Glu	Gly	Leu
				845					850					855
Ser	Val	Leu	Arg	Ser	Phe	Arg	Leu	Leu	Arg	Val	Phe	Lys	Leu	Ala
				860					865					870
Lys	Ser	Trp	Pro	Thr	Leu	Asn	Met	Leu	Ile	Lys	Ile	Ile	Gly	Asn
				875					880					885
Ser	Gly	Gly	Ala	Leu	Gly	Asn	Leu	Thr	Leu	Val	Leu	Ala	Ile	Ile
				890					895					900
Val	Phe	Ile	Phe	Ala	Val	Val	Gly	Met	Gln	Leu	Phe	Gly	Lys	Ser
				905					910					915
Tyr	Lys	Asp	Cys	Val	Cys	Lys	Ile	Ala	Ser	Asp	Cys	Gln	Leu	Pro
				920					925					930
Arg	Trp	His	Met	Asn	Asp	Phe	Phe	His	Ser	Phe	Leu	Ile	Val	Phe
				935					940					945
Arg	Val	Leu	Cys	Gly	Glu	Trp	Ile	Glu	Thr	Met	Trp	Asp	Cys	Met
				950					955					960
Glu	Val	Ala	Gly	Gln	Ala	Met	Cys	Leu	Thr	Val	Phe	Met	Met	Val
				965					970					975
Met	Val	Ile	Gly	Asn	Leu	Val	Val	Leu	Asn	Leu	Phe	Leu	Ala	Leu
				980					985					990
Leu	Leu	Ser	Ser	Phe	Ser	Ala	Asp	Asn	Leu	Ala	Ala	Thr	Asp	Asp
				995					1000					1005
Asp	Asn	Glu	Met	Asn	Asn	Leu	Gln	Ile	Ala	Val	Asp	Arg	Met	His
				1010					1015					1020
Lys	Gly	Val	Ala	Tyr	Val	Lys	Arg	Lys	Ile	Tyr	Glu	Phe	Ile	Gln
				1025					1030					1035
Gln	Ser	Phe	Ile	Arg	Lys	Gln	Lys	Ile	Leu	Asp	Glu	Ile	Lys	Pro
				1040					1045					1050
Leu	Asp	Asp	Leu	Asn	Asn	Lys	Lys	Asp	Ser	Cys	Met	Ser	Asn	His
				1055					1060					1065
Thr	Ala	Glu	Ile	Gly	Lys	Asp	Leu	Asp	Tyr	Leu	Lys	Asp	Val	Asn
				1070					1075					1080
Gly	Thr	Thr	Ser	Gly	Ile	Gly	Thr	Gly	Ser	Ser	Val	Glu	Lys	Tyr
				1085					1090					1095
Ile	Ile	Asp	Glu	Ser	Asp	Tyr	Met	Ser	Phe	Ile	Asn	Asn	Pro	Ser
				1100					1105					1110
Leu	Thr	Val	Thr	Val	Pro	Ile	Ala	Val	Gly	Glu	Ser	Asp	Phe	Glu
				1115					1120					1125
Asn	Leu	Asn	Thr	Glu	Asp	Phe	Ser	Ser	Glu	Ser	Asp	Leu	Glu	Glu
				1130					1135					1140
Ser	Lys	Glu	Lys	Leu	Asn	Glu	Ser	Ser	Ser	Ser	Ser	Glu	Gly	Ser
				1145					1150					1155
Thr	Val	Asp	Ile	Gly	Ala	Pro	Val	Glu	Glu	Gln	Pro	Val	Val	Glu
				1160					1165					1170
Pro	Glu	Glu	Thr	Leu	Glu	Pro	Glu	Ala	Cys	Phe	Thr	Glu	Gly	Cys

1175	1180	1185
Val Gln Arg Phe Lys Cys Cys Gln Ile Asn	Val Glu Glu Gly Arg	
1190	1195	1200
Gly Lys Gln Trp Trp Asn Leu Arg Arg Thr	Cys Phe Arg Ile Val	
1205	1210	1215
Glu His Asn Trp Phe Glu Thr Phe Ile Val	Phe Met Ile Leu Leu	
1220	1225	1230
Ser Ser Gly Ala Leu Ala Phe Glu Asp Ile	Tyr Ile Asp Gln Arg	
1235	1240	1245
Lys Thr Ile Lys Thr Met Leu Glu Tyr Ala	Asp Lys Val Phe Thr	
1250	1255	1260
Tyr Ile Phe Ile Leu Glu Met Leu Leu Lys	Trp Val Ala Tyr Gly	
1265	1270	1275
Tyr Gln Thr Tyr Phe Thr Asn Ala Trp Cys	Trp Leu Asp Phe Leu	
1280	1285	1290
Ile Val Asp Val Ser Leu Val Ser Leu Thr	Ala Asn Ala Leu Gly	
1295	1300	1305
Tyr Ser Glu Leu Gly Ala Ile Lys Ser Leu	Arg Thr Leu Arg Ala	
1310	1315	1320
Leu Arg Pro Leu Arg Ala Leu Ser Arg Phe	Glu Gly Met Arg Val	
1325	1330	1335
Val Val Asn Ala Leu Leu Gly Ala Ile Pro	Ser Ile Met Asn Val	
1340	1345	1350
Leu Leu Val Cys Leu Ile Phe Trp Leu Ile	Phe Ser Ile Met Gly	
1355	1360	1365
Val Asn Leu Phe Ala Gly Lys Phe Tyr His	Cys Ile Asn Thr Thr	
1370	1375	1380
Thr Gly Asp Arg Phe Asp Ile Glu Asp Val	Asn Asn His Thr Asp	
1385	1390	1395
Cys Leu Lys Leu Ile Glu Arg Asn Glu Thr	Ala Arg Trp Lys Asn	
1400	1405	1410
Val Lys Val Asn Phe Asp Asn Val Gly Phe	Gly Tyr Leu Ser Leu	
1415	1420	1425
Leu Gln Val Ala Thr Phe Lys Gly Trp Met	Asp Ile Met Tyr Ala	
1430	1435	1440
Ala Val Asp Ser Arg Asn Val Glu Leu Gln	Pro Lys Tyr Glu Glu	
1445	1450	1455
Ser Leu Tyr Met Tyr Leu Tyr Phe Val Ile	Phe Ile Ile Phe Gly	
1460	1465	1470
Ser Phe Phe Thr Leu Asn Leu Phe Ile Gly	Val Ile Ile Asp Asn	
1475	1480	1485
Phe Asn Gln Gln Lys Lys Lys Phe Gly Gly	Gln Asp Ile Phe Met	
1490	1495	1500
Thr Glu Glu Gln Lys Lys Tyr Tyr Asn Ala	Met Lys Lys Leu Gly	
1505	1510	1515
Ser Lys Lys Pro Gln Lys Pro Ile Pro Arg	Pro Gly Asn Lys Phe	
1520	1525	1530
Gln Gly Met Val Phe Asp Phe Val Thr Arg	Gln Val Phe Asp Ile	
1535	1540	1545
Ser Ile Met Ile Leu Ile Cys Leu Asn Met	Val Thr Met Met Val	
1550	1555	1560
Glu Thr Asp Asp Gln Ser Glu Tyr Val Thr	Thr Ile Leu Ser Arg	
1565	1570	1575
Ile Asn Leu Val Phe Ile Val Leu Phe Thr	Gly Glu Cys Val Leu	
1580	1585	1590
Lys Leu Ile Ser Leu Arg His Tyr Tyr Phe	Thr Ile Gly Trp Asn	
1595	1600	1605
Ile Phe Asp Phe Val Val Val Ile Leu Ser	Ile Val Gly Met Phe	
1610	1615	1620
Leu Ala Glu Leu Ile Glu Lys Tyr Phe Val	Ser Pro Thr Leu Phe	
1625	1630	1635
Arg Val Ile Arg Leu Ala Arg Ile Gly Arg	Ile Leu Arg Leu Ile	
1640	1645	1650

Lys Gly Ala Lys Gly Ile Arg Thr Leu Leu Phe Ala Leu Met Met  
 1655 1660 1665  
 Ser Leu Pro Ala Leu Phe Asn Ile Gly Leu Leu Leu Phe Leu Val  
 1670 1675 1680  
 Met Phe Ile Tyr Ala Ile Phe Gly Met Ser Asn Phe Ala Tyr Val  
 1685 1690 1695  
 Lys Arg Glu Val Gly Ile Asp Asp Met Phe Asn Phe Glu Thr Phe  
 1700 1705 1710  
 Gly Asn Ser Met Ile Cys Leu Phe Gln Ile Thr Thr Ser Ala Gly  
 1715 1720 1725  
 Trp Asp Gly Leu Leu Ala Pro Ile Leu Asn Ser Lys Pro Pro Asp  
 1730 1735 1740  
 Cys Asp Pro Asn Lys Val Asn Pro Gly Ser Ser Val Lys Gly Asp  
 1745 1750 1755  
 Cys Gly Asn Pro Ser Val Gly Ile Phe Phe Phe Val Ser Tyr Ile  
 1760 1765 1770  
 Ile Ile Ser Phe Leu Val Val Val Asn Met Tyr Ile Ala Val Ile  
 1775 1780 1785  
 Leu Glu Asn Phe Ser Val Ala Thr Glu Glu Ser Ala Glu Pro Leu  
 1790 1795 1800  
 Ser Glu Asp Asp Phe Glu Met Phe Tyr Glu Val Trp Glu Lys Phe  
 1805 1810 1815  
 Asp Pro Asp Ala Thr Gln Phe Met Glu Phe Glu Lys Leu Ser Gln  
 1820 1825 1830  
 Phe Ala Ala Ala Leu Glu Pro Pro Leu Asn Leu Pro Gln Pro Asn  
 1835 1840 1845  
 Lys Leu Gln Leu Ile Ala Met Asp Leu Pro Met Val Ser Gly Asp  
 1850 1855 1860  
 Arg Ile His Cys Leu Asp Ile Leu Phe Ala Phe Thr Lys Arg Val  
 1865 1870 1875  
 Leu Gly Glu Ser Gly Glu Met Asp Ala Leu Arg Ile Gln Met Glu  
 1880 1885 1890  
 Glu Arg Phe Met Ala Ser Asn Pro Ser Lys Val Ser Tyr Gln Pro  
 1895 1900 1905  
 Ile Thr Thr Thr Lys Arg Lys Gln Glu Glu Val Ser Ala Val  
 1910 1915 1920  
 Ile Ile Gln Arg Ala Tyr Arg Arg His Leu Leu Lys Arg Thr Val  
 1925 1930 1935  
 Lys Gln Ala Ser Phe Thr Tyr Asn Lys Asn Lys Ile Lys Gly Gly  
 1940 1945 1950  
 Ala Asn Leu Leu Ile Lys Glu Asp Met Ile Ile Asp Arg Ile Asn  
 1955 1960 1965  
 Glu Asn Ser Ile Thr Glu Lys Thr Asp Leu Thr Met Ser Thr Ala  
 1970 1975 1980  
 Ala Cys Pro Pro Ser Tyr Asp Arg Val Thr Lys Pro Ile Val Glu  
 1985 1990 1995  
 Lys His Glu Gln Glu Gly Lys Asp Glu Lys Ala Lys Gly Lys  
 2000 2005

&lt;210&gt; 14

&lt;211&gt; 538

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 70035348CD1

&lt;400&gt; 14

Met Val Pro Val Glu Asn Thr Glu Gly Pro Ser Leu Leu Asn Gln  
 1 5 10 15  
 Lys Gly Thr Ala Val Glu Thr Glu Gly Ser Gly Ser Arg His Pro  
 20 25 30



Pro	Trp	Ala	Arg	Gly	Cys	Gly	Met	Phe	Thr	Phe	Leu	Ser	Ser	Val
				35					40					45
Thr	Ala	Ala	Val	Ser	Gly	Leu	Leu	Val	Gly	Tyr	Glu	Leu	Gly	Ile
				50					55					60
Ile	Ser	Gly	Ala	Leu	Leu	Gln	Ile	Lys	Thr	Leu	Leu	Ala	Leu	Ser
				65					70					75
Cys	His	Glu	Gln	Glu	Met	Val	Val	Ser	Ser	Leu	Val	Ile	Gly	Ala
				80					85					90
Leu	Leu	Ala	Ser	Leu	Thr	Gly	Gly	Val	Leu	Ile	Asp	Arg	Tyr	Gly
				95					100					105
Arg	Arg	Thr	Ala	Ile	Ile	Leu	Ser	Ser	Cys	Leu	Leu	Gly	Leu	Gly
				110					115					120
Ser	Leu	Val	Leu	Ile	Leu	Ser	Leu	Ser	Tyr	Thr	Val	Leu	Ile	Val
				125					130					135
Gly	Arg	Ile	Ala	Ile	Gly	Val	Ser	Ile	Ser	Leu	Ser	Ser	Ile	Ala
				140					145					150
Thr	Cys	Val	Tyr	Ile	Ala	Glu	Ile	Ala	Pro	Gln	His	Arg	Arg	Gly
				155					160					165
Leu	Leu	Val	Ser	Leu	Asn	Glu	Leu	Met	Ile	Val	Ile	Gly	Ile	Leu
				170					175					180
Ser	Ala	Tyr	Ile	Ser	Asn	Tyr	Ala	Phe	Ala	Asn	Val	Phe	His	Gly
				185					190					195
Trp	Lys	Tyr	Met	Phe	Gly	Leu	Val	Ile	Pro	Leu	Gly	Val	Leu	Gln
				200					205					210
Ala	Ile	Ala	Met	Tyr	Phe	Leu	Pro	Pro	Ser	Pro	Arg	Phe	Leu	Val
				215					220					225
Met	Lys	Gly	Gln	Glu	Gly	Ala	Ala	Ser	Lys	Val	Leu	Gly	Arg	Leu
				230					235					240
Arg	Ala	Leu	Ser	Asp	Thr	Thr	Glu	Glu	Leu	Thr	Val	Ile	Lys	Ser
				245					250					255
Ser	Leu	Lys	Asp	Glu	Tyr	Gln	Tyr	Ser	Phe	Trp	Asp	Leu	Phe	Arg
				260					265					270
Ser	Lys	Asp	Asn	Met	Arg	Thr	Arg	Ile	Met	Ile	Gly	Leu	Thr	Leu
				275					280					285
Val	Phe	Phe	Val	Gln	Ile	Thr	Gly	Gln	Pro	Asn	Ile	Leu	Phe	Tyr
				290					295					300
Ala	Ser	Thr	Val	Leu	Lys	Ser	Val	Gly	Phe	Gln	Ser	Asn	Glu	Ala
				305					310					315
Ala	Ser	Leu	Ala	Ser	Thr	Gly	Val	Gly	Val	Val	Lys	Val	Ile	Ser
				320					325					330
Thr	Ile	Pro	Ala	Thr	Leu	Leu	Val	Asp	His	Val	Gly	Ser	Lys	Thr
				335					340					345
Phe	Leu	Cys	Ile	Gly	Ser	Ser	Val	Met	Ala	Ala	Ser	Leu	Val	Thr
				350					355					360
Met	Gly	Ile	Val	Asn	Leu	Asn	Ile	His	Met	Asn	Phe	Thr	His	Ile
				365					370					375
Cys	Arg	Ser	His	Asn	Ser	Ile	Asn	Gln	Ser	Leu	Asp	Glu	Ser	Val
				380					385					390
Ile	Tyr	Gly	Pro	Gly	Asn	Leu	Ser	Thr	Asn	Asn	Asn	Thr	Leu	Arg
				395					400					405
Asp	His	Phe	Lys	Gly	Ile	Ser	Ser	His	Ser	Arg	Ser	Ser	Leu	Met
				410					415					420
Pro	Leu	Arg	Asn	Asp	Val	Asp	Lys	Arg	Gly	Glu	Thr	Thr	Ser	Ala
				425					430					435
Ser	Leu	Leu	Asn	Ala	Gly	Leu	Ser	His	Thr	Glu	Tyr	Gln	Ile	Val
				440					445					450
Thr	Asp	Pro	Gly	Asp	Val	Pro	Ala	Phe	Leu	Lys	Trp	Leu	Ser	Leu
				455					460					465
Ala	Ser	Leu	Leu	Val	Tyr	Val	Ala	Ala	Phe	Ser	Ile	Gly	Leu	Gly
				470					475					480
Pro	Arg	Asp	Val	Ile	Phe	Ile	Gly	Gln	Ser	Thr	Asn	Leu	Pro	Ser
				485					490					495
Ala	Pro	Glu	Gly	Asp	Thr	Ile	Ser	Ile	Ser	Lys	Thr	Ile	Tyr	Tyr

Ala Ala Tyr Asn	500	Lys Ala Ile Ile Gln	505	Thr Ala Leu Glu Arg	510
	515		520		525
Pro Arg Ala Lys		Thr Val Ser Ala Phe		Ser His Lys Thr	
	530		535		

&lt;210&gt; 15

&lt;211&gt; 742

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7472539CD1

&lt;400&gt; 15

Met Glu Tyr Gln Ala	1	Ser Glu Val Ile Gly	10	Gln Arg Gln Ser Ser	15
Ala Thr Lys Pro Gly	20	Arg Ser Gly Lys Glu	25	Ser Val Thr Glu Pro	30
Trp Ala Arg Val Pro	35	Gly Ala Leu Gly Val	40	Ala Ala Arg Gln Met	45
His Pro Lys Ser Ile	50	Ile Thr Phe Arg Glu	55	Ile Asn Gly Glu Tyr	60
Thr Gly Ala Val Asp	65	Phe Pro Arg Leu Gly	70	Val Arg Ala Ser Glu	75
Glu Thr Ala Leu Arg	80	Glu Leu Lys Met Ser	85	Lys Glu Leu Ala Ala	90
Met Gly Pro Gly Ala	95	Ser Gly Asp Gly Val	100	Arg Thr Glu Thr Ala	105
Pro His Ile Ala Leu	110	Asp Ser Arg Val Gly	115	Leu His Ala Tyr Asp	120
Ile Ser Val Val Val	125	Ile Tyr Phe Val Phe	130	Val Ile Ala Val Gly	135
Ile Trp Ser Ser Ile	140	Arg Ala Ser Arg Gly	145	Thr Ile Gly Gly Tyr	150
Phe Leu Ala Gly Ser	155	Trp Ser Ile Ser Asp	160	Val Gln Gln Cys Gly	165
Gln Trp Leu Val His	170	Arg Pro Gly Trp Asp	175	Arg Gly Cys Arg Arg	180
Pro Cys Arg Arg Trp	185	Leu Arg Val Glu Leu	190	Leu Leu Ala Leu Gly	195
Trp Val Phe Val Pro	200	Val Tyr Ile Ala Ala	205	Gly Val Val Thr Met	210
Pro Gln Tyr Leu Lys	215	Lys Arg Phe Gly Gly	220	Gln Arg Ile Gln Val	225
Tyr Met Ser Val Leu	230	Ser Leu Ile Leu Tyr	235	Ile Phe Thr Lys Ile	240
Ser Thr Asp Ile Phe	245	Ser Gly Ala Leu Phe	250	Ile Gln Met Ala Leu	255
Gly Trp Asn Leu Tyr	260	Leu Ser Thr Gly Ile	265	Leu Leu Val Val Thr	270
Ala Val Tyr Thr Ile	275	Ala Gly Gly Leu Met	280	Ala Val Ile Tyr Thr	285
Asp Ala Leu Gln Thr	290	Val Ile Met Val Gly	295	Gly Ala Leu Val Leu	300
Met Phe Leu Gly Phe	305	Gln Asp Val Gly Trp	310	Tyr Pro Gly Leu Glu	315
Gln Arg Tyr Arg Gln	320	Ala Ile Pro Asn Val	325	Thr Val Pro Asn Thr	330
Thr Cys His Leu Pro	335	Arg Pro Asp Ala Phe	340	His Ile Leu Arg Asp	345
Pro Val Ser Gly Asp		Ile Pro Trp Pro Gly		Leu Ile Phe Gly Leu	

Thr Val Leu Ala	350	Thr Trp Cys Trp Cys	355	Thr Asp Gln Val Ile	360
Gln Arg Ser Leu	365	Ser Ala Lys Ser Leu	370	Ser His Ala Lys Gly	375
Ser Val Leu Gly	380	Gly Tyr Leu Lys Ile	385	Leu Pro Met Phe Phe	390
Val Met Pro Gly	395	Met Ile Ser Arg Ala	400	Leu Phe Pro Asp Glu	405
Gly Cys Val Asp	410	Pro Asp Val Cys Gln	415	Arg Ile Cys Gly Ala	420
Val Gly Cys Ser	425	Asn Ile Ala Tyr Pro	430	Leu Val Met Ala	435
Met Pro Val Gly	440	Leu Arg Gly Leu Met	445	Ile Ala Val Ile Met	450
Ala Leu Met Ser	455	Ser Leu Thr Ser Ile	460	Phe Asn Ser Ser Ser	465
Leu Phe Thr Ile	470	Asp Val Trp Gln Arg	475	Phe Arg Arg Lys Ser	480
Glu Gln Glu Leu	485	Met Val Val Gly Arg	490	Val Phe Val Val Phe	495
Val Val Ile Ser	500	Ile Leu Trp Ile Pro	505	Ile Ile Gln Ser Ser	510
Ser Gly Gln Leu	515	Phe Asp Tyr Ile Gln	520	Ala Val Thr Ser Tyr	525
Ala Pro Pro Ile	530	Thr Ala Leu Phe Leu	535	Leu Ala Ile Phe Cys	540
Arg Val Thr Glu	545	Pro Gly Ala Phe Trp	550	Gly Leu Val Phe Gly	555
Gly Val Gly Leu	560	Leu Arg Met Ile Leu	565	Glu Phe Ser Tyr Pro	570
Pro Ala Cys Gly	575	Glu Val Asp Arg Arg	580	Pro Ala Val Leu Lys	585
Phe His Tyr Leu	590	Tyr Phe Ala Ile Leu	595	Leu Cys Gly Leu Thr	600
Ile Val Ile Val	605	Ile Leu Thr Arg Leu	610	Thr Trp Trp Thr Arg	615
Cys Pro Leu Ser	620	Glu Leu Glu Lys Glu	625	Ala His Glu Ser Thr	630
Glu Ile Ser Glu	635	Arg Pro Ala Gly Glu	640	Cys Pro Ala Gly Gly	645
Ala Ala Glu Asn	650	Ser Ser Leu Gly Gln	655	Glu Gln Pro Glu Ala	660
Ser Arg Ser Trp	665	Gly Lys Leu Leu Trp	670	Ser Trp Phe Cys Gly	675
Ser Gly Thr Pro	680	Glu Gln Ala Leu Ser	685	Pro Ala Glu Lys Ala	690
Leu Glu Gln Lys	695	Leu Thr Ser Ile Glu	700	Glu Glu Pro Leu Trp	705
His Val Cys Asn	710	Ile Asn Ala Val Leu	715	Leu Leu Ala Ile Asn	720
Phe Leu Trp Gly	725	Tyr Phe Ala	730		735
	740				

&lt;210&gt; 16

&lt;211&gt; 426

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 817477CD1

&lt;400&gt; 16

Met	Ala	Arg	Arg	Thr	Glu	Pro	Pro	Asp	Gly	Gly	Trp	Gly	Trp	Val
1				5					10					15
Val	Val	Leu	Ser	Ala	Phe	Phe	Gln	Ser	Ala	Leu	Val	Phe	Gly	Val
				20					25					30
Leu	Arg	Ser	Phe	Gly	Val	Phe	Phe	Val	Glu	Phe	Val	Ala	Ala	Phe
				35					40					45
Glu	Glu	Gln	Ala	Ala	Arg	Val	Ser	Trp	Ile	Ala	Ser	Ile	Gly	Ile
				50					55					60
Ala	Val	Gln	Gln	Phe	Gly	Ser	Pro	Val	Gly	Ser	Ala	Leu	Ser	Thr
				65					70					75
Lys	Phe	Gly	Pro	Arg	Pro	Val	Val	Met	Thr	Gly	Gly	Ile	Leu	Ala
				80					85					90
Ala	Leu	Gly	Met	Leu	Leu	Ala	Ser	Phe	Ala	Thr	Ser	Leu	Thr	His
				95					100					105
Leu	Tyr	Leu	Ser	Ile	Gly	Leu	Leu	Ser	Gly	Ser	Gly	Trp	Ala	Leu
				110					115					120
Thr	Phe	Ala	Pro	Thr	Leu	Ala	Cys	Leu	Ser	Cys	Tyr	Phe	Ser	Arg
				125					130					135
Arg	Arg	Ser	Leu	Ala	Thr	Gly	Leu	Ala	Leu	Thr	Gly	Val	Gly	Leu
				140					145					150
Ser	Ser	Phe	Thr	Phe	Ala	Pro	Phe	Phe	Gln	Trp	Leu	Leu	Ser	His
				155					160					165
Tyr	Ala	Trp	Arg	Gly	Ser	Leu	Leu	Leu	Val	Ser	Ala	Leu	Ser	Leu
				170					175					180
His	Leu	Val	Ala	Cys	Gly	Ala	Leu	Leu	Arg	Pro	Pro	Ser	Leu	Ala
				185					190					195
Glu	Asp	Pro	Ala	Val	Gly	Gly	Pro	Arg	Ala	Gln	Leu	Thr	Ser	Leu
				200					205					210
Leu	His	His	Gly	Pro	Phe	Leu	Arg	Tyr	Thr	Val	Ala	Leu	Thr	Leu
				215					220					225
Ile	Asn	Thr	Gly	Tyr	Phe	Ile	Pro	Tyr	Leu	His	Leu	Val	Ala	His
				230					235					240
Leu	Gln	Asp	Leu	Asp	Trp	Asp	Pro	Leu	Pro	Ala	Ala	Phe	Leu	Leu
				245					250					255
Ser	Val	Val	Ala	Ile	Ser	Asp	Leu	Val	Gly	Arg	Val	Val	Ser	Gly
				260					265					270
Trp	Leu	Gly	Asp	Ala	Val	Pro	Gly	Pro	Val	Thr	Arg	Leu	Leu	Met
				275					280					285
Leu	Trp	Thr	Thr	Leu	Thr	Gly	Val	Ser	Leu	Ala	Leu	Phe	Pro	Val
				290					295					300
Ala	Gln	Ala	Pro	Thr	Ala	Leu	Val	Ala	Leu	Ala	Val	Ala	Tyr	Gly
				305					310					315
Phe	Thr	Ser	Gly	Ala	Leu	Ala	Pro	Leu	Ala	Phe	Ser	Val	Leu	Pro
				320					325					330
Glu	Leu	Ile	Gly	Thr	Arg	Arg	Ile	Tyr	Cys	Gly	Leu	Gly	Leu	Leu
				335					340					345
Gln	Met	Ile	Glu	Ser	Ile	Gly	Gly	Leu	Leu	Gly	Pro	Pro	Leu	Ser
				350					355					360
Gly	Tyr	Leu	Arg	Asp	Val	Thr	Gly	Asn	Tyr	Thr	Ala	Ser	Phe	Val
				365					370					375
Val	Ala	Gly	Ala	Phe	Leu	Leu	Ser	Gly	Ser	Gly	Ile	Leu	Leu	Thr
				380					385					390
Leu	Pro	His	Phe	Cys	Phe	Ser	Thr	Thr	Thr	Ser	Gly	Pro	Gln	
				395					400					405
Asp	Leu	Val	Thr	Glu	Ala	Leu	Asp	Thr	Lys	Val	Pro	Leu	Pro	Lys
				410					415					420
Glu	Gly	Leu	Glu	Glu	Asp									
				425										

&lt;210&gt; 17

&lt;211&gt; 1197

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1442166CD1

&lt;400&gt; 17

Met	Ala	Ala	Ala	Ala	Ala	Val	Gly	Asn	Ala	Val	Pro	Cys	Gly	Ala
1				5					10					15
Arg	Pro	Cys	Gly	Val	Arg	Pro	Asp	Gly	Gln	Pro	Lys	Pro	Gly	Pro
				20					25					30
Gln	Pro	Arg	Ala	Leu	Leu	Ala	Ala	Gly	Pro	Ala	Leu	Ile	Ala	Asn
				35					40					45
Gly	Asp	Glu	Leu	Val	Ala	Ala	Val	Trp	Pro	Tyr	Arg	Arg	Leu	Ala
				50					55					60
Leu	Leu	Arg	Arg	Leu	Thr	Val	Leu	Pro	Phe	Ala	Gly	Leu	Leu	Tyr
				65					70					75
Pro	Ala	Trp	Leu	Gly	Ala	Ala	Ala	Ala	Gly	Cys	Trp	Gly	Trp	Gly
				80					85					90
Ser	Ser	Trp	Val	Gln	Ile	Pro	Glu	Ala	Ala	Leu	Leu	Val	Leu	Ala
				95					100					105
Thr	Ile	Cys	Leu	Ala	His	Ala	Leu	Thr	Val	Leu	Ser	Gly	His	Trp
				110					115					120
Ser	Val	His	Ala	His	Cys	Ala	Leu	Thr	Cys	Thr	Pro	Glu	Tyr	Asp
				125					130					135
Pro	Ser	Lys	Ala	Thr	Phe	Val	Lys	Val	Val	Pro	Thr	Pro	Asn	Asn
				140					145					150
Gly	Ser	Thr	Glu	Leu	Val	Ala	Leu	His	Arg	Asn	Glu	Gly	Glu	Asp
				155					160					165
Gly	Leu	Glu	Val	Leu	Ser	Phe	Glu	Phe	Gln	Lys	Ile	Lys	Tyr	Ser
				170					175					180
Tyr	Asp	Ala	Leu	Glu	Lys	Lys	Gln	Phe	Leu	Pro	Val	Ala	Phe	Pro
				185					190					195
Val	Gly	Asn	Ala	Phe	Ser	Tyr	Tyr	Gln	Ser	Asn	Arg	Gly	Phe	Gln
				200					205					210
Glu	Asp	Ser	Glu	Ile	Arg	Ala	Ala	Glu	Lys	Lys	Phe	Gly	Ser	Asn
				215					220					225
Lys	Ala	Glu	Met	Val	Val	Pro	Asp	Phe	Ser	Glu	Leu	Phe	Lys	Glu
				230					235					240
Arg	Ala	Thr	Ala	Pro	Phe	Phe	Val	Phe	Gln	Val	Phe	Cys	Val	Gly
				245					250					255
Leu	Trp	Cys	Leu	Asp	Glu	Tyr	Trp	Tyr	Tyr	Ser	Val	Phe	Thr	Leu
				260					265					270
Ser	Met	Leu	Val	Ala	Phe	Glu	Ala	Ser	Leu	Val	Gln	Gln	Gln	Met
				275					280					285
Arg	Asn	Met	Ser	Glu	Ile	Arg	Lys	Met	Gly	Asn	Lys	Pro	His	Met
				290					295					300
Ile	Gln	Val	Tyr	Arg	Ser	Arg	Lys	Trp	Arg	Pro	Ile	Ala	Ser	Asp
				305					310					315
Glu	Ile	Val	Pro	Gly	Asp	Ile	Val	Ser	Ile	Gly	Arg	Ser	Pro	Gln
				320					325					330
Glu	Asn	Leu	Val	Pro	Cys	Asp	Val	Leu	Leu	Leu	Arg	Gly	Arg	Cys
				335					340					345
Ile	Val	Asp	Glu	Ala	Met	Leu	Thr	Gly	Glu	Ser	Val	Pro	Gln	Met
				350					355					360
Lys	Glu	Pro	Ile	Glu	Asp	Leu	Ser	Pro	Asp	Arg	Val	Leu	Asp	Leu
				365					370					375
Gln	Ala	Asp	Ser	Arg	Leu	His	Val	Ile	Phe	Gly	Gly	Thr	Lys	Val
				380					385					390
Val	Gln	His	Ile	Pro	Pro	Gln	Lys	Ala	Thr	Thr	Gly	Leu	Lys	Pro
				395					400					405
Val	Asp	Ser	Gly	Cys	Val	Ala	Tyr	Val	Leu	Arg	Thr	Gly	Phe	Asn
				410					415					420

Thr	Ser	Gln	Gly	Lys	Leu	Leu	Arg	Thr	Ile	Leu	Phe	Gly	Val	Lys
				425					430					435
Arg	Val	Thr	Ala	Asn	Asn	Leu	Glu	Thr	Phe	Ile	Phe	Ile	Leu	Phe
				440					445					450
Leu	Leu	Val	Phe	Ala	Ile	Ala	Ala	Ala	Ala	Tyr	Val	Trp	Ile	Glu
				455					460					465
Gly	Thr	Lys	Asp	Pro	Ser	Arg	Asn	Arg	Tyr	Lys	Leu	Phe	Leu	Glu
				470					475					480
Cys	Thr	Leu	Ile	Leu	Thr	Ser	Val	Val	Pro	Pro	Glu	Leu	Pro	Ile
				485					490					495
Glu	Leu	Ser	Leu	Ala	Val	Asn	Thr	Ser	Leu	Ile	Ala	Leu	Ala	Lys
				500					505					510
Leu	Tyr	Met	Tyr	Cys	Thr	Glu	Pro	Phe	Arg	Ile	Pro	Phe	Ala	Gly
				515					520					525
Lys	Val	Glu	Val	Cys	Cys	Phe	Asp	Lys	Thr	Gly	Thr	Leu	Thr	Ser
				530					535					540
Asp	Ser	Leu	Val	Val	Arg	Gly	Val	Ala	Gly	Leu	Arg	Asp	Gly	Lys
				545					550					555
Glu	Val	Thr	Pro	Val	Ser	Ser	Ile	Pro	Val	Glu	Thr	His	Arg	Ala
				560					565					570
Leu	Ala	Ser	Cys	His	Ser	Leu	Met	Gln	Leu	Asp	Asp	Gly	Thr	Leu
				575					580					585
Val	Gly	Asp	Pro	Leu	Glu	Lys	Ala	Met	Leu	Thr	Ala	Val	Asp	Trp
				590					595					600
Thr	Leu	Thr	Lys	Asp	Glu	Lys	Val	Phe	Pro	Arg	Ser	Ile	Lys	Thr
				605					610					615
Gln	Gly	Leu	Lys	Ile	His	Gln	Arg	Phe	His	Phe	Ala	Ser	Ala	Leu
				620					625					630
Lys	Arg	Met	Ser	Val	Leu	Ala	Ser	Tyr	Glu	Lys	Leu	Gly	Ser	Thr
				635					640					645
Asp	Leu	Cys	Tyr	Ile	Ala	Ala	Val	Lys	Gly	Ala	Pro	Glu	Thr	Leu
				650					655					660
His	Ser	Met	Phe	Ser	Gln	Cys	Pro	Pro	Asp	Tyr	His	His	Ile	His
				665					670					675
Thr	Glu	Ile	Ser	Arg	Glu	Gly	Ala	Arg	Val	Leu	Ala	Leu	Gly	Tyr
				680					685					690
Lys	Glu	Leu	Gly	His	Leu	Thr	His	Gln	Gln	Ala	Arg	Glu	Val	Lys
				695					700					705
Arg	Glu	Ala	Leu	Glu	Cys	Ser	Leu	Lys	Phe	Val	Gly	Phe	Ile	Val
				710					715					720
Val	Ser	Cys	Pro	Leu	Lys	Ala	Asp	Ser	Lys	Ala	Val	Ile	Arg	Glu
				725					730					735
Ile	Gln	Asn	Ala	Ser	His	Arg	Val	Val	Met	Ile	Thr	Gly	Asp	Asn
				740					745					750
Pro	Leu	Thr	Ala	Cys	His	Val	Ala	Gln	Glu	Leu	His	Phe	Ile	Glu
				755					760					765
Lys	Ala	His	Thr	Leu	Ile	Leu	Gln	Pro	Pro	Ser	Glu	Lys	Gly	Arg
				770					775					780
Gln	Cys	Glu	Trp	Arg	Ser	Ile	Asp	Gly	Ser	Ile	Val	Leu	Pro	Leu
				785					790					795
Ala	Arg	Gly	Ser	Pro	Lys	Ala	Leu	Ala	Leu	Glu	Tyr	Ala	Leu	Cys
				800					805					810
Leu	Thr	Gly	Asp	Gly	Leu	Ala	His	Leu	Gln	Ala	Thr	Asp	Pro	Gln
				815					820					825
Gln	Leu	Leu	Arg	Leu	Ile	Pro	His	Val	Gln	Val	Phe	Ala	Arg	Val
				830					835					840
Ala	Pro	Lys	Gln	Lys	Glu	Phe	Val	Ile	Thr	Ser	Leu	Lys	Glu	Leu
				845					850					855
Gly	Tyr	Val	Thr	Leu	Met	Cys	Gly	Asp	Gly	Thr	Asn	Asp	Val	Gly
				860					865					870
Ala	Leu	Lys	His	Ala	Asp	Val	Gly	Val	Ala	Leu	Leu	Ala	Asn	Ala
				875					880					885
Pro	Glu	Arg	Val	Val	Glu	Arg	Arg	Arg	Arg	Pro	Arg	Asp	Ser	Pro

				890					895					900
Thr	Leu	Ser	Asn	Ser	Gly	Ile	Arg	Ala	Thr	Ser	Arg	Thr	Ala	Lys
				905					910					915
Gln	Arg	Ser	Gly	Leu	Pro	Pro	Ser	Glu	Glu	Gln	Pro	Thr	Ser	Gln
				920					925					930
Arg	Asp	Arg	Leu	Ser	Gln	Val	Leu	Arg	Asp	Leu	Glu	Asp	Glu	Ser
				935					940					945
Thr	Pro	Ile	Val	Lys	Leu	Gly	Asp	Ala	Ser	Ile	Ala	Ala	Pro	Phe
				950					955					960
Thr	Ser	Lys	Leu	Ser	Ser	Ile	Gln	Cys	Ile	Cys	His	Val	Ile	Lys
				965					970					975
Gln	Gly	Arg	Cys	Thr	Leu	Val	Thr	Thr	Leu	Gln	Met	Phe	Lys	Ile
				980					985					990
Leu	Ala	Leu	Asn	Ala	Leu	Ile	Leu	Ala	Tyr	Ser	Gln	Ser	Val	Leu
				995					1000					1005
Tyr	Leu	Glu	Gly	Val	Lys	Phe	Ser	Asp	Phe	Gln	Ala	Thr	Leu	Gln
				1010					1015					1020
Gly	Leu	Leu	Leu	Ala	Gly	Cys	Phe	Leu	Phe	Ile	Ser	Arg	Ser	Lys
				1025					1030					1035
Pro	Leu	Lys	Thr	Leu	Ser	Arg	Glu	Arg	Pro	Leu	Pro	Asn	Ile	Phe
				1040					1045					1050
Asn	Leu	Tyr	Thr	Ile	Leu	Thr	Val	Met	Leu	Gln	Phe	Phe	Val	His
				1055					1060					1065
Phe	Leu	Ser	Leu	Val	Tyr	Leu	Tyr	Arg	Glu	Ala	Gln	Ala	Arg	Ser
				1070					1075					1080
Pro	Glu	Lys	Gln	Glu	Gln	Phe	Val	Asp	Leu	Tyr	Lys	Glu	Phe	Glu
				1085					1090					1095
Pro	Ser	Leu	Val	Asn	Ser	Thr	Val	Tyr	Ile	Met	Ala	Met	Ala	Met
				1100					1105					1110
Gln	Met	Ala	Thr	Phe	Ala	Ile	Asn	Tyr	Lys	Gly	Pro	Pro	Phe	Met
				1115					1120					1125
Glu	Ser	Leu	Pro	Glu	Asn	Lys	Pro	Leu	Val	Trp	Ser	Leu	Ala	Val
				1130					1135					1140
Ser	Leu	Leu	Ala	Ile	Ile	Gly	Leu	Leu	Leu	Gly	Ser	Ser	Pro	Asp
				1145					1150					1155
Phe	Asn	Ser	Gln	Phe	Gly	Leu	Val	Asp	Ile	Pro	Val	Glu	Val	Leu
				1160					1165					1170
Leu	Leu	Asp	Phe	Cys	Leu	Ala	Leu	Leu	Ala	Asp	Arg	Val	Leu	Gln
				1175					1180					1185
Phe	Phe	Leu	Gly	Thr	Pro	Lys	Leu	Lys	Val	Pro	Ser			
				1190					1195					

&lt;210&gt; 18

&lt;211&gt; 1771

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2311751CD1

&lt;400&gt; 18

Met	Met	Glu	Arg	Ala	Ile	Ile	Asp	Thr	Phe	Val	Gly	His	Asp	Val
1				5					10					15
Val	Glu	Pro	Gly	Ser	Tyr	Val	Gln	Met	Phe	Pro	Tyr	Pro	Cys	Tyr
				20					25					30
Thr	Arg	Asp	Asp	Phe	Leu	Phe	Val	Ile	Glu	His	Met	Met	Pro	Leu
				35					40					45
Cys	Met	Val	Ile	Ser	Trp	Val	Tyr	Ser	Val	Ala	Met	Thr	Ile	Gln
				50					55					60
His	Ile	Val	Ala	Glu	Lys	Glu	His	Arg	Leu	Lys	Glu	Val	Met	Lys
				65					70					75
Thr	Met	Gly	Leu	Asn	Asn	Ala	Val	His	Trp	Val	Ala	Trp	Phe	Ile

	80		85		90
Thr Gly Phe Val	Gln Leu Ser Ile Ser	Val Thr Ala Leu Thr	Ala		
	95	100	105		
Ile Leu Lys Tyr	Gly Gln Val Leu Met	His Ser His Val Val	Ile		
	110	115	120		
Ile Trp Leu Phe	Leu Ala Val Tyr Ala	Val Ala Thr Ile Met	Phe		
	125	130	135		
Cys Phe Leu Val	Ser Val Leu Tyr Ser	Lys Ala Lys Leu Ala	Ser		
	140	145	150		
Ala Cys Gly Gly	Ile Ile Tyr Phe Leu	Ser Tyr Val Pro Tyr	Met		
	155	160	165		
Tyr Val Ala Ile	Arg Glu Glu Val Ala	His Asp Lys Ile Thr	Ala		
	170	175	180		
Phe Glu Lys Cys	Ile Ala Ser Leu Met	Ser Thr Thr Ala Phe	Gly		
	185	190	195		
Leu Gly Ser Lys	Tyr Phe Ala Leu Tyr	Glu Val Ala Gly Val	Gly		
	200	205	210		
Ile Gln Trp His	Thr Phe Ser Gln Ser	Pro Val Glu Gly Asp	Asp		
	215	220	225		
Phe Asn Leu Leu	Leu Ala Val Thr Met	Leu Met Val Asp Ala	Val		
	230	235	240		
Val Tyr Gly Ile	Leu Thr Trp Tyr Ile	Glu Ala Val His Pro	Gly		
	245	250	255		
Met Tyr Gly Leu	Pro Arg Pro Trp Tyr	Phe Pro Leu Gln Lys	Ser		
	260	265	270		
Tyr Trp Leu Gly	Ser Gly Arg Thr Glu	Ala Trp Glu Trp Ser	Trp		
	275	280	285		
Pro Trp Ala Arg	Thr Pro Arg Leu Ser	Val Met Glu Glu Asp	Gln		
	290	295	300		
Ala Cys Ala Met	Glu Ser Arg Arg Phe	Glu Glu Thr Arg Gly	Met		
	305	310	315		
Glu Glu Glu Pro	Thr His Leu Pro Leu	Val Val Cys Val Asp	Lys		
	320	325	330		
Leu Thr Lys Val	Tyr Lys Asp Asp Lys	Lys Leu Ala Leu Asn	Lys		
	335	340	345		
Leu Ser Leu Asn	Leu Tyr Glu Asn Gln	Val Val Ser Phe Leu	Gly		
	350	355	360		
His Asn Gly Ala	Gly Lys Thr Thr Thr	Met Ser Ile Leu Thr	Gly		
	365	370	375		
Leu Phe Pro Pro	Thr Ser Gly Ser Ala	Thr Ile Tyr Gly His	Asp		
	380	385	390		
Ile Arg Thr Glu	Met Asp Glu Ile Arg	Lys Asn Leu Gly Met	Cys		
	395	400	405		
Pro Gln His Asn	Val Leu Phe Asp Arg	Leu Thr Val Glu Glu	His		
	410	415	420		
Leu Trp Phe Tyr	Ser Arg Leu Lys Ser	Met Ala Gln Glu Glu	Ile		
	425	430	435		
Arg Arg Glu Met	Asp Lys Met Ile Glu	Asp Leu Glu Leu Ser	Asn		
	440	445	450		
Lys Arg His Ser	Leu Val Gln Thr Leu	Ser Gly Gly Met Lys	Arg		
	455	460	465		
Lys Leu Ser Val	Ala Ile Ala Phe Val	Gly Gly Ser Arg Ala	Ile		
	470	475	480		
Ile Leu Asp Glu	Pro Thr Ala Gly Val	Asp Pro Tyr Ala Arg	Arg		
	485	490	495		
Ala Ile Trp Asp	Leu Ile Leu Lys Tyr	Lys Pro Gly Arg Thr	Ile		
	500	505	510		
Leu Leu Ser Thr	His His Met Asp Glu	Ala Asp Leu Leu Gly	Asp		
	515	520	525		
Arg Ile Ala Ile	Ile Ser His Gly Lys	Leu Lys Cys Cys Gly	Ser		
	530	535	540		
Pro Leu Phe Leu	Lys Gly Thr Tyr Gly	Asp Gly Tyr Arg Leu	Thr		
	545	550	555		



Leu	Val	Lys	Arg	Pro	Ala	Glu	Pro	Gly	Gly	Pro	Gln	Glu	Pro	Gly
				560					565					570
Leu	Ala	Ser	Ser	Pro	Pro	Gly	Arg	Ala	Pro	Leu	Ser	Ser	Cys	Ser
				575					580					585
Glu	Leu	Gln	Val	Ser	Gln	Phe	Ile	Arg	Lys	His	Val	Ala	Ser	Cys
				590					595					600
Leu	Leu	Val	Ser	Asp	Thr	Ser	Thr	Glu	Leu	Ser	Tyr	Ile	Leu	Pro
				605					610					615
Ser	Glu	Ala	Ala	Lys	Lys	Gly	Ala	Phe	Glu	Arg	Leu	Phe	Gln	His
				620					625					630
Leu	Glu	Arg	Ser	Leu	Asp	Ala	Leu	His	Leu	Ser	Ser	Phe	Gly	Leu
				635					640					645
Met	Asp	Thr	Thr	Leu	Glu	Glu	Val	Phe	Leu	Lys	Val	Ser	Glu	Glu
				650					655					660
Asp	Gln	Ser	Leu	Glu	Asn	Ser	Glu	Ala	Asp	Val	Lys	Glu	Ser	Arg
				665					670					675
Lys	Asp	Val	Leu	Pro	Gly	Ala	Glu	Gly	Pro	Ala	Ser	Gly	Glu	Gly
				680					685					690
His	Ala	Gly	Asn	Leu	Ala	Arg	Cys	Ser	Glu	Leu	Thr	Gln	Ser	Gln
				695					700					705
Ala	Ser	Leu	Gln	Ser	Ala	Ser	Ser	Val	Gly	Ser	Ala	Arg	Gly	Asp
				710					715					720
Glu	Gly	Ala	Gly	Tyr	Thr	Asp	Val	Tyr	Gly	Asp	Tyr	Arg	Pro	Leu
				725					730					735
Phe	Asp	Asn	Pro	Gln	Asp	Pro	Asp	Asn	Val	Ser	Leu	Gln	Glu	Val
				740					745					750
Glu	Ala	Glu	Ala	Leu	Ser	Arg	Val	Gly	Gln	Gly	Ser	Arg	Lys	Leu
				755					760					765
Asp	Gly	Gly	Trp	Leu	Lys	Val	Arg	Gln	Phe	His	Gly	Leu	Leu	Val
				770					775					780
Lys	Arg	Phe	His	Cys	Ala	Arg	Arg	Asn	Ser	Lys	Ala	Leu	Phe	Ser
				785					790					795
Gln	Ile	Leu	Leu	Pro	Ala	Phe	Phe	Val	Cys	Val	Ala	Met	Thr	Val
				800					805					810
Ala	Leu	Ser	Val	Pro	Glu	Ile	Gly	Asp	Leu	Pro	Pro	Leu	Val	Leu
				815					820					825
Ser	Pro	Ser	Gln	Tyr	His	Asn	Tyr	Thr	Gln	Pro	Arg	Gly	Asn	Phe
				830					835					840
Ile	Pro	Tyr	Ala	Asn	Glu	Glu	Arg	Arg	Glu	Tyr	Arg	Leu	Arg	Leu
				845					850					855
Ser	Pro	Asp	Ala	Ser	Pro	Gln	Gln	Leu	Val	Ser	Thr	Phe	Arg	Leu
				860					865					870
Pro	Ser	Gly	Val	Gly	Ala	Thr	Cys	Val	Leu	Lys	Ser	Pro	Ala	Asn
				875					880					885
Gly	Ser	Leu	Gly	Pro	Thr	Leu	Asn	Leu	Ser	Ser	Gly	Glu	Ser	Arg
				890					895					900
Leu	Leu	Ala	Ala	Arg	Phe	Phe	Asp	Ser	Met	Cys	Leu	Glu	Ser	Phe
				905					910					915
Thr	Gln	Gly	Leu	Pro	Leu	Ser	Asn	Phe	Val	Pro	Pro	Pro	Pro	Ser
				920					925					930
Pro	Ala	Pro	Ser	Asp	Ser	Pro	Ala	Ser	Pro	Asp	Glu	Asp	Leu	Gln
				935					940					945
Ala	Trp	Asn	Val	Ser	Leu	Pro	Pro	Thr	Ala	Gly	Pro	Glu	Met	Trp
				950					955					960
Thr	Ser	Ala	Pro	Ser	Leu	Pro	Arg	Leu	Val	Arg	Glu	Pro	Val	Arg
				965					970					975
Cys	Thr	Cys	Ser	Ala	Gln	Gly	Thr	Gly	Phe	Ser	Cys	Pro	Ser	Ser
				980					985					990
Val	Gly	Gly	His	Pro	Pro	Gln	Met	Arg	Val	Val	Thr	Gly	Asp	Ile
				995					1000					1005
Leu	Thr	Asp	Ile	Thr	Gly	His	Asn	Val	Ser	Glu	Tyr	Leu	Leu	Phe
				1010					1015					1020
Thr	Ser	Asp	Arg	Phe	Arg	Leu	His	Arg	Tyr	Gly	Ala	Ile	Thr	Phe

1025	1030	1035
Gly Asn Val Leu Lys Ser Ile Pro Ala Ser	Phe Gly Thr Arg Ala	
1040	1045	1050
Pro Pro Met Val Arg Lys Ile Ala Val Arg	Arg Ala Ala Gln Val	
1055	1060	1065
Phe Tyr Asn Asn Lys Gly Tyr His Ser Met	Pro Thr Tyr Leu Asn	
1070	1075	1080
Ser Leu Asn Asn Ala Ile Leu Arg Ala Asn	Leu Pro Lys Ser Lys	
1085	1090	1095
Gly Asn Pro Ala Ala Tyr Gly Ile Thr Val	Thr Asn His Pro Met	
1100	1105	1110
Asn Lys Thr Ser Ala Ser Leu Ser Leu Asp	Tyr Leu Leu Gln Gly	
1115	1120	1125
Thr Asp Val Val Ile Ala Ile Phe Ile Ile	Val Ala Met Ser Phe	
1130	1135	1140
Val Pro Ala Ser Phe Val Val Phe Leu Val	Ala Glu Lys Ser Thr	
1145	1150	1155
Lys Ala Lys His Leu Gln Phe Val Ser Gly	Cys Asn Pro Ile Ile	
1160	1165	1170
Tyr Trp Leu Ala Asn Tyr Val Trp Asp Met	Leu Asn Tyr Leu Val	
1175	1180	1185
Pro Ala Thr Cys Cys Val Ile Ile Leu Phe	Val Phe Asp Leu Pro	
1190	1195	1200
Ala Tyr Thr Ser Pro Thr Asn Phe Pro Ala	Val Leu Ser Leu Phe	
1205	1210	1215
Leu Leu Tyr Gly Trp Ser Ile Thr Pro Ile	Met Tyr Pro Ala Ser	
1220	1225	1230
Phe Trp Phe Glu Val Pro Ser Ser Ala Tyr	Val Phe Leu Ile Val	
1235	1240	1245
Ile Asn Leu Phe Ile Gly Ile Thr Ala Thr	Val Ala Thr Phe Leu	
1250	1255	1260
Leu Gln Leu Phe Glu His Asp Lys Asp Leu	Lys Val Val Asn Ser	
1265	1270	1275
Tyr Leu Lys Ser Cys Phe Leu Ile Phe Pro	Asn Tyr Asn Leu Gly	
1280	1285	1290
His Gly Leu Met Glu Met Ala Tyr Asn Glu	Tyr Ile Asn Glu Tyr	
1295	1300	1305
Tyr Ala Lys Ile Gly Gln Phe Asp Lys Met	Lys Ser Pro Phe Glu	
1310	1315	1320
Trp Asp Ile Val Thr Arg Gly Leu Val Ala	Met Ala Val Glu Gly	
1325	1330	1335
Val Val Gly Phe Leu Leu Thr Ile Met Cys	Gln Tyr Asn Phe Leu	
1340	1345	1350
Arg Arg Pro Gln Arg Met Pro Val Ser Thr	Lys Pro Val Glu Asp	
1355	1360	1365
Asp Val Asp Val Ala Ser Glu Arg Gln Arg	Val Leu Arg Gly Asp	
1370	1375	1380
Ala Asp Asn Asp Met Val Lys Ile Glu Asn	Leu Thr Lys Val Tyr	
1385	1390	1395
Lys Ser Arg Lys Ile Gly Arg Ile Leu Ala	Val Asp Arg Leu Cys	
1400	1405	1410
Leu Gly Val Arg Pro Gly Glu Cys Phe Gly	Leu Leu Gly Val Asn	
1415	1420	1425
Gly Ala Gly Lys Thr Ser Thr Phe Lys Met	Leu Thr Gly Asp Glu	
1430	1435	1440
Ser Thr Thr Gly Gly Glu Ala Phe Val Asn	Gly His Ser Val Leu	
1445	1450	1455
Lys Glu Leu Leu Gln Val Gln Gln Ser Leu	Gly Tyr Cys Pro Gln	
1460	1465	1470
Cys Asp Ala Leu Phe Asp Glu Leu Thr Ala	Arg Glu His Leu Gln	
1475	1480	1485
Leu Tyr Thr Arg Leu Arg Gly Ile Ser Trp	Lys Asp Glu Ala Arg	
1490	1495	1500

```
<210> 19
<211> 474
<212> PRT
<213> Homo sapiens
```

```
<220>  
<221> misc_feature  
<223> Incyte ID No: 7472537CD1
```

<400> 19														
Met	Phe	Ser	Leu	Ser	Tyr	Leu	Cys	Val	Cys	Val	Phe	Ser	Gln	Phe
1				5					10					15
Ala	Asn	Glu	Asp	Thr	Glu	Ser	Gln	Lys	Phe	Leu	Thr	Asn	Gly	Phe
				20					25					30
Leu	Gly	Lys	Lys	Lys	Leu	Ala	Asp	Pro	Phe	Phe	Phe	Lys	His	Pro
				35					40					45
Gly	Thr	Thr	Ser	Phe	Gly	Met	Ser	Ser	Phe	Asn	Leu	Ser	Asn	Ala
				50					55					60
Ile	Met	Gly	Ser	Gly	Ile	Leu	Gly	Leu	Ser	Tyr	Ala	Met	Ala	Asn
				65					70					75
Thr	Gly	Ile	Ile	Leu	Phe	Met	Phe	Met	Leu	Leu	Ala	Val	Ala	Ile
				80					85					90
Leu	Ser	Leu	Tyr	Ser	Val	His	Leu	Leu	Leu	Lys	Thr	Ser	Leu	Ile
				95					100					105

Val	Gly	Ser	Leu	Ile	Tyr	Glu	Lys	Leu	Gly	Glu	Lys	Ala	Phe	Gly	
				110					115					120	
Trp	Pro	Gly	Lys	Ile	Gly	Ala	Phe	Val	Ser	Ile	Thr	Met	Gln	Asn	
				125					130					135	
Ile	Gly	Ala	Met	Ser	Ser	Tyr	Leu	Phe	Ile	Ile	Lys	Tyr	Glu	Leu	
				140					145					150	
Pro	Glu	Val	Ile	Arg	Ala	Phe	Met	Gly	Leu	Glu	Glu	Thr	Ser	Arg	
				155					160					165	
Glu	Trp	Tyr	Leu	Asn	Gly	Asn	Tyr	Leu	Ile	Ile	Phe	Val	Ser	Val	
				170					175					180	
Gly	Ile	Ile	Leu	Pro	Leu	Ser	Leu	Leu	Lys	Asn	Leu	Gly	Tyr	Leu	
				185					190					195	
Gly	Tyr	Thr	Ser	Gly	Phe	Ser	Leu	Thr	Cys	Met	Val	Phe	Phe	Val	
				200					205					210	
Ser	Val	Val	Ile	Tyr	Lys	Lys	Phe	Gln	Ile	Pro	Cys	Pro	Leu	Pro	
				215					220					225	
Glu	Asn	Gln	Ala	Lys	Gly	Ser	Leu	His	Asp	Ser	Gly	Val	Glu	Tyr	
				230					235					240	
Glu	Ala	His	Ser	Asp	Asp	Lys	Cys	Glu	Pro	Lys	Tyr	Phe	Val	Phe	
				245					250					255	
Asn	Ser	Gln	Thr	Ala	Tyr	Ala	Ile	Pro	Ile	Leu	Val	Phe	Ala	Phe	
				260					265					270	
Val	Cys	His	Pro	Glu	Val	Leu	Pro	Ile	Tyr	Ser	Glu	Leu	Lys	Asp	
				275					280					285	
Arg	Ser	Arg	Arg	Lys	Met	Gln	Thr	Val	Ser	Asn	Ile	Ser	Ile	Thr	
				290					295					300	
Gly	Met	Leu	Val	Met	Tyr	Leu	Leu	Ala	Ala	Leu	Phe	Gly	Tyr	Leu	
				305					310					315	
Thr	Phe	Tyr	Gly	Arg	Val	Glu	Asp	Glu	Leu	Leu	His	Ala	Tyr	Ser	
				320					325					330	
Lys	Val	Tyr	Thr	Leu	Asp	Ile	Pro	Leu	Leu	Met	Val	Arg	Leu	Ala	
				335					340					345	
Val	Leu	Val	Ala	Val	Thr	Leu	Thr	Val	Pro	Ile	Val	Leu	Phe	Pro	
				350					355					360	
Val	Arg	Thr	Ser	Val	Ile	Thr	Leu	Leu	Phe	Pro	Lys	Arg	Pro	Phe	
				365					370					375	
Ser	Trp	Ile	Arg	His	Phe	Leu	Ile	Ala	Ala	Val	Leu	Ile	Ala	Leu	
				380					385					390	
Asn	Asn	Val	Leu	Val	Ile	Leu	Val	Pro	Thr	Ile	Lys	Tyr	Ile	Phe	
				395					400					405	
Gly	Phe	Ile	Gly	Ala	Ser	Ser	Ala	Thr	Met	Leu	Ile	Phe	Ile	Leu	
				410					415					420	
Pro	Ala	Val	Phe	Tyr	Leu	Lys	Leu	Val	Lys	Lys	Glu	Thr	Phe	Arg	
				425					430					435	
Ser	Pro	Pro	Glu	Leu	Gln	Ala	Leu	Ile	Phe	Leu	Val	Val	Gly	Ile	
				440					445					450	
Phe	Phe	Met	Ile	Gly	Ser	Met	Ala	Leu	Ile	Ile	Ile	Asp	Trp	Ile	
				455					460					465	
Tyr	Asp	Pro	Pro	Asn	Ser	Lys	His	His							
				470											

&lt;210&gt; 20

&lt;211&gt; 752

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7472546CD1

&lt;400&gt; 20

Met	Glu	Tyr	Gln	Ala	Ser	Glu	Val	Ile	Gly	Gln	Arg	Gln	Ser	Ser	
1				5					10					15	

Ala	Thr	Lys	Pro	Gly	Arg	Ser	Gly	Lys	Glu	Ser	Val	Thr	Glu	Pro
				20					25					30
Trp	Ala	Arg	Val	Pro	Gly	Ala	Leu	Gly	Val	Ala	Ala	Arg	Gln	Met
				35					40					45
His	Pro	Lys	Ser	Ile	Ile	Thr	Phe	Arg	Glu	Ile	Asn	Gly	Glu	Tyr
				50					55					60
Thr	Gly	Ala	Val	Asp	Phe	Pro	Arg	Leu	Gly	Val	Arg	Ala	Ser	Glu
				65					70					75
Glu	Thr	Ala	Leu	Arg	Glu	Leu	Lys	Met	Ser	Lys	Glu	Leu	Ala	Ala
				80					85					90
Met	Gly	Pro	Gly	Ala	Ser	Gly	Asp	Gly	Val	Arg	Thr	Glu	Thr	Ala
				95					100					105
Pro	His	Ile	Ala	Leu	Asp	Ser	Arg	Val	Gly	Leu	His	Ala	Tyr	Asp
				110					115					120
Ile	Ser	Val	Val	Val	Ile	Tyr	Phe	Val	Phe	Val	Ile	Ala	Val	Gly
				125					130					135
Ile	Trp	Ser	Ser	Ile	Arg	Ala	Ser	Arg	Gly	Thr	Ile	Gly	Gly	Tyr
				140					145					150
Phe	Leu	Ala	Gly	Arg	Ser	Met	Ser	Trp	Trp	Pro	Ile	Gly	Ala	Ser
				155					160					165
Leu	Met	Ser	Ser	Asn	Val	Gly	Ser	Gly	Leu	Phe	Ile	Gly	Leu	Ala
				170					175					180
Gly	Thr	Gly	Ala	Ala	Gly	Gly	Leu	Ala	Val	Gly	Gly	Phe	Glu	Trp
				185					190					195
Asn	Ala	Thr	Trp	Leu	Leu	Leu	Ala	Leu	Gly	Trp	Val	Phe	Val	Pro
				200					205					210
Val	Tyr	Ile	Ala	Ala	Gly	Val	Val	Thr	Met	Pro	Gln	Tyr	Leu	Lys
				215					220					225
Lys	Arg	Phe	Gly	Gly	Gln	Arg	Ile	Gln	Val	Tyr	Met	Ser	Val	Leu
				230					235					240
Ser	Leu	Ile	Leu	Tyr	Ile	Phe	Thr	Lys	Ile	Ser	Thr	Asp	Ile	Phe
				245					250					255
Ser	Gly	Ala	Leu	Phe	Ile	Gln	Met	Ala	Leu	Gly	Trp	Asn	Leu	Tyr
				260					265					270
Leu	Ser	Thr	Gly	Ile	Leu	Leu	Val	Val	Thr	Ala	Val	Tyr	Thr	Ile
				275					280					285
Ala	Gly	Gly	Leu	Met	Ala	Val	Ile	Tyr	Thr	Asp	Ala	Leu	Gln	Thr
				290					295					300
Val	Ile	Met	Val	Gly	Gly	Ala	Leu	Val	Leu	Met	Phe	Leu	Gly	Phe
				305					310					315
Gln	Asp	Val	Gly	Trp	Tyr	Pro	Gly	Leu	Glu	Gln	Arg	Tyr	Arg	Gln
				320					325					330
Ala	Ile	Pro	Asn	Val	Thr	Val	Pro	Asn	Thr	Thr	Cys	His	Leu	Pro
				335					340					345
Arg	Pro	Asp	Ala	Phe	His	Ile	Leu	Arg	Asp	Pro	Val	Ser	Gly	Asp
				350					355					360
Ile	Pro	Trp	Pro	Gly	Leu	Ile	Phe	Gly	Leu	Thr	Val	Leu	Ala	Thr
				365					370					375
Trp	Cys	Trp	Cys	Thr	Asp	Gln	Val	Ile	Val	Gln	Arg	Ser	Leu	Ser
				380					385					390
Ala	Lys	Ser	Leu	Ser	His	Ala	Lys	Gly	Gly	Ser	Val	Leu	Gly	Gly
				395					400					405
Tyr	Leu	Lys	Ile	Leu	Pro	Met	Phe	Phe	Ile	Val	Met	Pro	Gly	Met
				410					415					420
Ile	Ser	Arg	Ala	Leu	Phe	Pro	Asp	Glu	Val	Gly	Cys	Val	Asp	Pro
				425					430					435
Asp	Val	Cys	Gln	Arg	Ile	Cys	Gly	Ala	Arg	Val	Gly	Cys	Ser	Asn
				440					445					450
Ile	Ala	Tyr	Pro	Lys	Leu	Val	Met	Ala	Leu	Met	Pro	Val	Gly	Leu
				455					460					465
Arg	Gly	Leu	Met	Ile	Ala	Val	Ile	Met	Ala	Ala	Leu	Met	Ser	Ser
				470					475					480
Leu	Thr	Ser	Ile	Phe	Asn	Ser	Ser	Ser	Thr	Leu	Phe	Thr	Ile	Asp

	485		490		495
Val Trp Gln Arg	Phe Arg Arg Lys Ser	Thr Glu Gln Glu Leu	Met		
	500		505		510
Val Val Gly Arg	Val Phe Val Val Phe	Leu Val Val Ile Ser	Ile		
	515		520		525
Leu Trp Ile Pro	Ile Ile Gln Ser Ser	Asn Ser Gly Gln Leu	Phe		
	530		535		540
Asp Tyr Ile Gln	Ala Val Thr Ser Tyr	Leu Ala Pro Pro Ile	Thr		
	545		550		555
Ala Leu Phe Leu	Leu Ala Ile Phe Cys	Lys Arg Val Thr Glu	Pro		
	560		565		570
Gly Ala Phe Trp	Gly Leu Val Phe Gly	Leu Gly Val Gly Leu	Leu		
	575		580		585
Arg Met Ile Leu	Glu Phe Ser Tyr Pro	Ala Pro Ala Cys Gly	Glu		
	590		595		600
Val Asp Arg Arg	Pro Ala Val Leu Lys	Asp Phe His Tyr Leu	Tyr		
	605		610		615
Phe Ala Ile Leu	Leu Cys Gly Leu Thr	Ala Ile Val Ile Val	Ile		
	620		625		630
Leu Thr Arg Leu	Thr Trp Trp Thr Arg	Asn Cys Pro Leu Ser	Glu		
	635		640		645
Leu Glu Lys Glu	Ala His Glu Ser Thr	Pro Glu Ile Ser Glu	Arg		
	650		655		660
Pro Ala Gly Glu	Cys Pro Ala Gly Gly	Gly Ala Ala Glu Asn	Ser		
	665		670		675
Ser Leu Gly Gln	Glu Gln Pro Glu Ala	Pro Ser Arg Ser Trp	Gly		
	680		685		690
Lys Leu Leu Trp	Ser Trp Phe Cys Gly	Leu Ser Gly Thr Pro	Glu		
	695		700		705
Gln Ala Leu Ser	Pro Ala Glu Lys Ala	Ala Leu Glu Gln Lys	Leu		
	710		715		720
Thr Ser Ile Glu	Glu Glu Pro Leu Trp	Arg His Val Cys Asn	Ile		
	725		730		735
Asn Ala Val Leu	Leu Leu Ala Ile Asn	Ile Phe Leu Trp Gly	Tyr		
	740		745		750
Phe Ala					

<210> 21  
 <211> 654  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7474202CD1

<400> 21  
 Met Glu Glu Leu Val Gly Leu Arg Glu Gly Phe Ser Gly Asp Pro  
 1 5 10 15  
 Val Thr Leu Gln Glu Leu Trp Gly Pro Cys Pro His Ile Arg Arg  
 20 25 30  
 Ala Ile Gln Gly Gly Leu Glu Trp Leu Lys Gln Lys Val Phe Arg  
 35 40 45  
 Leu Gly Glu Asp Trp Tyr Phe Leu Met Thr Leu Gly Val Leu Met  
 50 55 60  
 Ala Leu Val Ser Tyr Ala Met Asn Phe Ala Ile Gly Cys Val Val  
 65 70 75  
 Arg Gly Phe Ser Gln Ser Ile Thr Pro Ser Ser Gly Gly Ser Gly  
 80 85 90  
 Ile Pro Glu Leu Lys Thr Met Leu Ala Gly Val Ile Leu Glu Asp  
 95 100 105  
 Tyr Leu Asp Ile Lys Asn Phe Gly Ala Lys Val Val Gly Leu Ser

	110		115		120
Cys Thr Leu Ala	Thr Gly Ser Thr Leu	Phe Leu Gly Lys Val	Gly		
	125		130		135
Pro Phe Val His	Leu Ser Val Met Ile	Ala Ala Tyr Leu Gly	Arg		
	140		145		150
Val Arg Thr Thr	Thr Ile Gly Glu Pro	Glu Asn Lys Ser Lys	Gln		
	155		160		165
Asn Glu Met Leu	Val Ala Ala Ala Ala	Val Gly Val Ala Thr	Val		
	170		175		180
Phe Ala Ala Pro	Phe Ser Gly Val Leu	Phe Ser Ile Glu Val	Met		
	185		190		195
Ser Ser His Phe	Ser Val Arg Asp Tyr	Trp Arg Gly Phe Phe	Ala		
	200		205		210
Ala Thr Cys Gly	Ala Phe Ile Phe Arg	Leu Leu Ala Val Phe	Asn		
	215		220		225
Ser Glu Gln Glu	Thr Ile Thr Ser Leu	Tyr Lys Thr Ser Phe	Arg		
	230		235		240
Val Asp Val Pro	Phe Asp Leu Pro Glu	Ile Phe Phe Phe Val	Ala		
	245		250		255
Leu Gly Gly Ile	Cys Gly Val Leu Ser	Cys Ala Tyr Leu Phe	Cys		
	260		265		270
Gln Arg Thr Phe	Leu Ser Phe Ile Lys	Thr Asn Arg Tyr Ser	Ser		
	275		280		285
Lys Leu Leu Ala	Thr Ser Lys Pro Val	Tyr Ser Ala Leu Ala	Thr		
	290		295		300
Leu Leu Leu Ala	Ser Ile Thr Tyr Pro	Pro Gly Val Gly His	Phe		
	305		310		315
Leu Ala Ser Arg	Leu Ser Met Lys Gln	His Leu Asp Ser Leu	Phe		
	320		325		330
Asp Asn His Ser	Trp Ala Leu Met Thr	Gln Asn Ser Ser Pro	Pro		
	335		340		345
Trp Pro Glu Glu	Leu Asp Pro Gln His	Leu Trp Trp Glu Trp	Tyr		
	350		355		360
His Pro Arg Phe	Thr Ile Phe Gly Thr	Leu Ala Phe Phe Leu	Val		
	365		370		375
Met Lys Phe Trp	Met Leu Ile Leu Ala	Thr Thr Ile Pro Met	Pro		
	380		385		390
Ala Gly Tyr Phe	Met Pro Ile Phe Ile	Leu Gly Ala Ala Ile	Gly		
	395		400		405
Arg Leu Leu Gly	Glu Ala Leu Ala Val	Ala Phe Pro Glu Gly	Ile		
	410		415		420
Val Thr Gly Gly	Val Thr Asn Pro Ile	Met Pro Gly Gly Tyr	Ala		
	425		430		435
Leu Ala Gly Ala	Ala Ala Phe Ser Gly	Ala Val Thr His Thr	Ile		
	440		445		450
Ser Thr Ala Leu	Leu Ala Phe Glu Leu	Thr Gly Gln Ile Val	His		
	455		460		465
Ala Leu Pro Val	Leu Met Ala Val Leu	Ala Ala Asn Ala Ile	Ala		
	470		475		480
Gln Ser Cys Gln	Pro Ser Phe Tyr Asp	Gly Thr Ile Ile Val	Lys		
	485		490		495
Lys Leu Pro Tyr	Leu Pro Arg Ile Leu	Gly Arg Asn Ile Gly	Ser		
	500		505		510
His His Val Arg	Val Glu His Phe Met	Asn His Ser Ile Thr	Thr		
	515		520		525
Leu Ala Lys Asp	Thr Pro Leu Glu Glu	Val Val Lys Val Val	Thr		
	530		535		540
Ser Thr Asp Val	Thr Glu Tyr Pro Leu	Val Glu Ser Thr Glu	Ser		
	545		550		555
Gln Ile Leu Val	Gly Ile Val Gln Arg	Ala Gln Leu Val Gln	Ala		
	560		565		570
Leu Gln Ala Glu	Pro Pro Ser Arg Ala	Pro Gly His Gln Cys	Leu		
	575		580		585

Gln Asp Ile Leu Ala Arg Gly Cys Pro Thr Glu Pro Val Thr Leu  
 590 595 600  
 Thr Leu Phe Ser Glu Thr Thr Leu His Gln Ala Gln Asn Leu Phe  
 605 610 615  
 Lys Leu Leu Asn Leu Gln Ser Leu Phe Val Thr Ser Arg Gly Arg  
 620 625 630  
 Ala Val Gly Cys Val Ser Trp Val Glu Met Lys Lys Ala Ile Ser  
 635 640 645  
 Asn Leu Thr Asn Pro Pro Ala Pro Lys  
 650

<210> 22

<211> 886

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7476280CD1

<400> 22

Met Asp Pro Ile Thr Pro Asn Trp Thr Glu Ile Val Asn Arg Lys  
 1 5 10 15  
 Leu Ser Phe Pro Pro Pro Leu Leu Asp Ala Ile Gln Glu Gly Arg  
 20 25 30  
 Leu Gly Phe Val Gln Gln Leu Leu Glu Ser Glu Val Glu Ala Ala  
 35 40 45  
 Ser Ser Gly Pro Gly Trp Pro Leu Trp Asn Val Glu Glu Ala Glu  
 50 55 60  
 Asp Arg Cys Trp Arg Glu Ala Leu Asn Leu Ala Ile Arg Leu Gly  
 65 70 75  
 His Glu Ala Leu Thr Asp Val Leu Leu Ala Ser Val Lys Phe Asp  
 80 85 90  
 Phe Arg Gln Ile His Glu Ala Leu Leu Val Ala Val Asp Thr Asn  
 95 100 105  
 Gln Ala Val Val Arg Arg Leu Pro Ala Arg Leu Glu Arg Glu Lys  
 110 115 120  
 Gly Arg Lys Val Asp Thr Arg Ser Phe Ser Leu Ala Phe Phe Asp  
 125 130 135  
 Ser Ser Ile Asp Gly Ser Arg Phe Ala Pro Gly Val Thr Pro Leu  
 140 145 150  
 Pro Gln Ala Cys Gln Lys Asp Leu Tyr Glu Ile Ala Gln Leu Leu  
 155 160 165  
 Met Glu Gln Gly His Thr Ile Ala Arg Pro His Pro Val Ser Cys  
 170 175 180  
 Ala Cys Leu Glu Cys Ser Asn Ala Arg Arg Tyr Asp Leu Leu Lys  
 185 190 195  
 Leu Ser Leu Ser Arg Ile Asn Thr Tyr Leu Gly Ile Ala Ser Arg  
 200 205 210  
 Ala His Leu Ser Leu Ala Ser Glu Asp Ala Met Leu Ala Ala Phe  
 215 220 225  
 Gln Leu Ser Arg Glu Leu Arg Arg Leu Ala Arg Lys Glu Pro Glu  
 230 235 240  
 Phe Lys Pro Glu Tyr Ile Ala Leu Glu Ser Leu Ser Gln Asp Tyr  
 245 250 255  
 Gly Phe Gln Leu Leu Gly Met Cys Trp Asn Gln Ser Glu Val Thr  
 260 265 270  
 Ala Val Leu Asn Asp Leu Ala Glu Asp Ser Glu Thr Glu Pro Glu  
 275 280 285  
 Ala Glu Gly Leu Gly Leu Ala Phe Glu Glu Gly Ile Pro Asn Leu  
 290 295 300  
 Val Arg Leu Arg Leu Ala Val Asn Tyr Asn Gln Lys Arg Phe Val  
 305 310 315



Ala His Leu Ile Cys Gln Gln Val Leu Ser Ser Ile Trp Cys Gly	320	325	330
Asn Leu Ala Gly Trp Arg Gly Ser Thr Thr Ser Trp Lys Leu Phe	335	340	345
Ala Thr Phe Leu Ile Phe Leu Thr Met Pro Phe Leu Cys Leu Gly	350	355	360
Tyr Trp Leu Thr Pro Lys Ser Gln Leu Gly His Leu Leu Lys Ile	365	370	375
Pro Val Leu Lys Phe Leu Leu His Ser Ala Ser Tyr Leu Trp Phe	380	385	390
Leu Ile Phe Leu Leu Gly Glu Ser Leu Val Met Glu Thr Gln Leu	395	400	405
Ser Thr Phe Arg Gly Arg Ser Gln Ser Val Trp Glu Thr Ser Leu	410	415	420
His Met Ile Cys Val Thr Gly Phe Leu Trp Phe Glu Cys Lys Glu	425	430	435
Val Trp Ile Glu Gly Leu Arg Ser Tyr Leu Leu Asp Trp Trp Asn	440	445	450
Phe Leu Asp Met Val Val Leu Ser Leu Tyr Leu Ala Ala Phe Ala	455	460	465
Leu Arg Leu Leu Leu Ala Gly Leu Ala Pro Met His Cys Arg Asp	470	475	480
Ala Ser Gln Ala Ala Ala Cys His Tyr Phe Thr Met Ala Glu Arg	485	490	495
Ser Glu Trp His Thr Glu Asp Pro Gln Phe Leu Ala Glu Val Leu	500	505	510
Phe Thr Ala Thr Ser Met Leu Ser Phe Thr Arg Leu Ala Tyr Ile	515	520	525
Leu Pro Ala His Glu Ser Leu Gly Thr Leu Gln Ile Ser Ile Gly	530	535	540
Lys Met Ile Glu Asp Met Ile Arg Phe Met Phe Ile Leu Met Ile	545	550	555
Ile Leu Thr Ala Phe Leu Cys Gly Leu Asn Asn Ile Tyr Val Pro	560	565	570
Tyr Gln Lys Thr Glu Trp Leu Gly Lys Ser Phe Asn Glu Thr Phe	575	580	585
Gln Phe Leu Phe Trp Thr Met Phe Gly Met Glu Glu His Ser Val	590	595	600
Val Asp Val Pro Gln Phe Leu Val Pro Glu Phe Ala Gly Arg Ala	605	610	615
Leu Tyr Gly Ile Phe Thr Ile Ile Met Val Ile Val Leu Leu Asn	620	625	630
Met Leu Ile Ala Met Ile Thr Asn Ser Phe Gln Lys Ile Glu Asp	635	640	645
Asp Ala Asp Val Glu Trp Thr Phe Ala Arg Ser Lys Leu Tyr Leu	650	655	660
Phe Tyr Phe Arg Glu Gly Leu Thr Leu Pro Val Pro Phe Asn Ile	665	670	675
Leu Pro Ser Ser Lys Ala Val Phe Tyr Leu Leu Arg Arg Ile Cys	680	685	690
Gln Phe Ile Cys Cys Cys Cys Ser Cys Cys Lys Thr Lys Lys Pro	695	700	705
Asp Tyr Pro Pro Ile Pro Thr Phe Val Asn Pro Arg Ala Gly Ala	710	715	720
Val Pro Gly Glu Gly Glu Arg Gly Ser Tyr Arg Leu His Val Ile	725	730	735
Lys Ala Leu Val Gln Arg Tyr Thr Glu Thr Ala Arg Arg Glu Phe	740	745	750
Glu Glu Thr Arg Arg Lys Asp Leu Gly Asn Arg Leu Thr Glu Leu	755	760	765
Thr Lys Thr Ile Ser Arg Leu Gln Ser Glu Val Ala Gly Val Arg	770	775	780
Arg Thr Leu Ala Glu Gly Gly Thr Pro Arg Pro Pro Asp Gly Ala			

	785		790		795
Ser Val Leu Ser	His Tyr Ile Thr Gln	Val His Asn Ser Phe	Gln		
	800		805		810
Asn Leu Gly Pro	Pro Ile Pro Glu Thr	Pro Glu Leu Thr Gly	Pro		
	815		820		825
Gly Ile Val Arg	Thr Gln Glu Ser Ser	Gly Thr Gly Leu Gln	Asp		
	830		835		840
Thr Gly Gly Val	Arg Thr Leu Ala Ser	Gly Glu Ser Gly Pro	Cys		
	845		850		855
Ser Pro Ala His	Val Leu Val His Arg	Glu Gln Glu Ala Glu	Gly		
	860		865		870
Ala Gly Asp Leu	Pro Gln Gly Glu Asp	Ser Gly Thr Glu Arg	Arg		
	875		880		885
Ser					

&lt;210&gt; 23

&lt;211&gt; 512

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1713377CD1

&lt;400&gt; 23

Met Ala Gly Gly Met	Ser Ala Glu Cys	Pro Glu Pro Gly Pro Gly	
1	5	10	15
Gly Leu Gln Gly Gln	Ser Pro Gly Pro Gly	Arg Gln Cys Pro Pro	
	20	25	30
Pro Ile Thr Pro Thr	Ser Trp Ser Leu	Pro Pro Trp Arg Ala Tyr	
	35	40	45
Val Ala Ala Ala Val	Leu Cys Tyr Ile	Asn Leu Leu Asn Tyr Met	
	50	55	60
Asn Trp Phe Ile Ile	Ala Gly Val Leu	Leu Asp Ile Gln Glu Val	
	65	70	75
Phe Gln Ile Ser Asp	Asn His Ala Gly	Leu Leu Gln Thr Val Phe	
	80	85	90
Val Ser Cys Leu Leu	Leu Ser Ala Pro	Val Phe Gly Tyr Leu Gly	
	95	100	105
Asp Arg His Ser Arg	Lys Ala Thr Met	Ser Phe Gly Ile Leu Leu	
	110	115	120
Trp Ser Gly Ala Gly	Leu Ser Ser Ser	Phe Ile Ser Pro Arg Tyr	
	125	130	135
Ser Trp Leu Phe Phe	Leu Ser Arg Gly	Ile Val Gly Thr Gly Ser	
	140	145	150
Ala Ser Tyr Ser Thr	Ile Ala Pro Thr	Val Leu Gly Asp Leu Phe	
	155	160	165
Val Arg Asp Gln Arg	Thr Arg Val Leu	Ala Val Phe Tyr Ile Phe	
	170	175	180
Ile Pro Val Gly Ser	Gly Leu Gly Tyr	Val Leu Gly Ser Ala Val	
	185	190	195
Thr Met Leu Thr Gly	Asn Trp Arg Trp	Ala Leu Arg Val Met Pro	
	200	205	210
Cys Leu Glu Ala Val	Ala Leu Ile Leu	Leu Ile Leu Leu Val Pro	
	215	220	225
Asp Pro Pro Arg Gly	Ala Ala Glu Thr	Gln Gly Glu Gly Ala Val	
	230	235	240
Gly Gly Phe Arg Ser	Ser Trp Cys Glu	Asp Val Arg Tyr Leu Gly	
	245	250	255
Lys Asn Trp Ser Phe	Val Trp Ser Thr	Leu Gly Val Thr Ala Met	
	260	265	270
Ala Phe Val Thr Gly	Ala Leu Gly Phe	Trp Ala Pro Lys Phe Leu	

Leu Glu Ala Arg	275	Val Val His Gly Leu	280	Gln Pro Pro Cys Phe	285
	290		295		300
Glu Pro Cys Ser	305	Asn Pro Asp Ser Leu	310	Phe Gly Ala Leu	315
Ile Met Thr Gly	320	Val Ile Gly Val Ile	325	Leu Gly Ala Glu Ala	330
Arg Arg Tyr Lys	335	Lys Val Ile Pro Gly	340	Ala Glu Pro Leu Ile	345
Ala Ser Ser Leu	350	Leu Ala Thr Ala Pro	355	Cys Leu Tyr Leu Ala	360
Val Leu Ala Pro	365	Thr Thr Leu Leu Ala	370	Ser Tyr Val Phe Leu	375
Leu Gly Glu Leu	380	Leu Leu Ser Cys Asn	385	Trp Ala Val Val Ala	390
Ile Leu Leu Ser	395	Val Val Val Pro Arg	400	Cys Arg Gly Thr Ala	405
Ala Leu Gln Ile	410	Thr Val Gly His Ile	415	Leu Gly Asp Ala Gly	420
Pro Tyr Leu Thr	425	Gly Leu Ile Ser Ser	430	Val Leu Arg Ala Arg	435
Pro Asp Ser Tyr	440	Leu Gln Arg Phe Arg	445	Ser Leu Gln Gln Ser	450
Leu Cys Cys Ala	455	Phe Val Ile Ala Leu	460	Gly Gly Cys Phe Leu	465
Leu Thr Ala Leu	470	Tyr Leu Glu Arg Asp	475	Glu Thr Arg Ala Trp	480
Pro Val Thr Gly	485	Thr Pro Asp Ser Asn	490	Asp Val Asp Ser Asn	495
Leu Glu Arg Gln	500	Gly Leu Leu Ser Gly	505	Ala Gly Ala Ser Thr	510
Glu Pro					

&lt;210&gt; 24

&lt;211&gt; 475

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 5842557CD1

&lt;400&gt; 24

Met Ile Pro Ala Tyr	5	Ser Lys Asn Arg Ala	10	Tyr Ala Ile Phe Phe	15
Ile Val Phe Thr Val	20	Ile Gly Ser Leu Phe	25	Leu Met Asn Leu Leu	30
Thr Ala Ile Ile Tyr	35	Ser Gln Phe Arg Gly	40	Tyr Leu Met Lys Ser	45
Leu Gln Thr Ser Leu	50	Phe Arg Arg Arg Leu	55	Gly Thr Arg Ala Ala	60
Phe Glu Val Leu Ser	65	Ser Met Val Gly Glu	70	Gly Gly Ala Phe Pro	75
Gln Ala Val Gly Val	80	Lys Pro Gln Asn Leu	85	Leu Gln Val Leu Gln	90
Lys Val Gln Leu Asp	95	Ser Ser His Lys Gln	100	Ala Met Met Glu Lys	105
Val Arg Ser Tyr Asp	110	Ser Val Leu Leu Ser	115	Ala Glu Glu Phe Gln	120
Lys Leu Phe Asn Glu	125	Leu Asp Arg Ser Val	130	Val Lys Glu His Pro	135
Pro Arg Pro Glu Tyr		Gln Ser Pro Phe Leu		Gln Ser Ala Gln Phe	

140	145	150
Leu Phe Gly His Tyr Tyr Phe Asp Tyr	Leu Gly Asn Leu Ile Ala	
155	160	165
Leu Ala Asn Leu Val Ser Ile Cys Val	Phe Leu Val Leu Asp Ala	
170	175	180
Asp Val Leu Pro Ala Glu Arg Asp Asp	Phe Ile Leu Gly Ile Leu	
185	190	195
Asn Cys Val Phe Ile Val Tyr Tyr Leu	Leu Glu Met Leu Leu Lys	
200	205	210
Val Phe Ala Leu Gly Leu Arg Gly Tyr	Leu Ser Tyr Pro Ser Asn	
215	220	225
Val Phe Asp Gly Leu Leu Thr Val Val	Leu Leu Val Leu Glu Ile	
230	235	240
Ser Thr Leu Ala Val Tyr Arg Leu Pro	His Pro Gly Trp Arg Pro	
245	250	255
Glu Met Val Gly Leu Leu Ser Leu Trp	Asp Met Thr Arg Met Leu	
260	265	270
Asn Met Leu Ile Val Phe Arg Phe Leu	Arg Ile Ile Pro Ser Met	
275	280	285
Lys Pro Met Ala Val Val Ala Ser Thr	Val Leu Gly Leu Val Gln	
290	295	300
Asn Met Arg Ala Phe Gly Gly Ile Leu	Val Val Val Tyr Tyr Val	
305	310	315
Phe Ala Ile Ile Gly Ile Asn Leu Phe	Arg Gly Val Ile Val Ala	
320	325	330
Leu Pro Gly Asn Ser Ser Leu Ala Pro	Ala Asn Gly Ser Ala Pro	
335	340	345
Cys Gly Ser Phe Glu Gln Leu Glu Tyr	Trp Ala Asn Asn Phe Asp	
350	355	360
Asp Phe Ala Ala Ala Leu Val Thr Leu	Trp Asn Leu Met Val Val	
365	370	375
Asn Asn Trp Gln Val Phe Leu Asp Ala	Tyr Arg Arg Tyr Ser Gly	
380	385	390
Pro Trp Ser Lys Ile Tyr Phe Val Leu	Trp Trp Leu Val Ser Ser	
395	400	405
Val Ile Trp Val Asn Leu Phe Leu Ala	Leu Ile Leu Glu Asn Phe	
410	415	420
Leu His Lys Trp Asp Pro Arg Ser His	Leu Gln Pro Leu Ala Gly	
425	430	435
Thr Pro Glu Ala Thr Tyr Gln Met Thr	Val Glu Leu Leu Phe Arg	
440	445	450
Asp Ile Leu Glu Glu Pro Glu Glu Asp	Glu Leu Thr Glu Arg Leu	
455	460	465
Ser Gln His Pro His Leu Trp Leu Cys	Arg	
470	475	

&lt;210&gt; 25

&lt;211&gt; 537

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7476643CD1

&lt;400&gt; 25

Met Ala Arg Lys Gln Asn Arg Asn Ser Lys Glu Leu Gly Leu Val	
1 5 10 15	
Pro Leu Thr Asp Asp Thr Ser His Ala Arg Pro Pro Gly Pro Gly	
20 25 30	
Arg Ala Leu Leu Glu Cys Asp His Leu Arg Ser Gly Val Pro Gly	
35 40 45	
Gly Arg Arg Arg Lys Asp Trp Ser Cys Ser Leu Leu Val Ala Ser	

	50		55		60									
Leu	Ala	Gly	Ala	Phe	Gly	Ser	Ser	Phe	Leu	Tyr	Gly	Tyr	Asn	Leu
	65		70		75									
Ser	Val	Val	Asn	Ala	Pro	Thr	Pro	Tyr	Ile	Lys	Ala	Phe	Tyr	Asn
	80		85		90									
Glu	Ser	Trp	Glu	Arg	Arg	His	Gly	Arg	Pro	Ile	Asp	Pro	Asp	Thr
	95		100		105									
Leu	Thr	Leu	Leu	Trp	Ser	Val	Thr	Val	Ser	Ile	Phe	Ala	Ile	Gly
	110		115		120									
Gly	Leu	Val	Gly	Thr	Leu	Ile	Val	Lys	Met	Ile	Gly	Lys	Val	Leu
	125		130		135									
Gly	Arg	Lys	His	Thr	Leu	Leu	Ala	Asn	Asn	Gly	Phe	Ala	Ile	Ser
	140		145		150									
Ala	Ala	Leu	Leu	Met	Ala	Cys	Ser	Leu	Gln	Ala	Gly	Ala	Phe	Glu
	155		160		165									
Met	Leu	Ile	Val	Gly	Arg	Phe	Ile	Met	Gly	Ile	Asp	Gly	Gly	Val
	170		175		180									
Ala	Leu	Ser	Val	Leu	Pro	Met	Tyr	Leu	Ser	Glu	Ile	Ser	Pro	Lys
	185		190		195									
Glu	Ile	Arg	Gly	Ser	Leu	Gly	Gln	Val	Thr	Ala	Ile	Phe	Ile	Cys
	200		205		210									
Ile	Gly	Val	Phe	Thr	Gly	Gln	Leu	Leu	Gly	Leu	Pro	Glu	Leu	Leu
	215		220		225									
Gly	Lys	Glu	Ser	Thr	Trp	Pro	Tyr	Leu	Phe	Gly	Val	Ile	Val	Val
	230		235		240									
Pro	Ala	Val	Val	Gln	Leu	Leu	Ser	Leu	Pro	Phe	Leu	Pro	Asp	Ser
	245		250		255									
Pro	Arg	Tyr	Leu	Leu	Leu	Glu	Lys	His	Asn	Glu	Ala	Arg	Ala	Val
	260		265		270									
Lys	Ala	Phe	Gln	Thr	Phe	Leu	Gly	Lys	Ala	Asp	Val	Ser	Gln	Glu
	275		280		285									
Val	Glu	Glu	Val	Leu	Ala	Glu	Ser	Arg	Val	Gln	Arg	Ser	Ile	Arg
	290		295		300									
Leu	Val	Ser	Val	Leu	Glu	Leu	Leu	Arg	Ala	Pro	Tyr	Val	Arg	Trp
	305		310		315									
Gln	Val	Val	Thr	Val	Ile	Val	Thr	Met	Ala	Cys	Tyr	Gln	Leu	Cys
	320		325		330									
Gly	Leu	Asn	Ala	Ile	Trp	Phe	Tyr	Thr	Asn	Ser	Ile	Phe	Gly	Lys
	335		340		345									
Ala	Gly	Ile	Pro	Leu	Ala	Lys	Ile	Pro	Tyr	Val	Thr	Leu	Ser	Thr
	350		355		360									
Gly	Gly	Ile	Glu	Thr	Leu	Ala	Ala	Val	Phe	Ser	Gly	Leu	Val	Ile
	365		370		375									
Glu	His	Leu	Gly	Arg	Arg	Pro	Leu	Leu	Ile	Gly	Gly	Phe	Gly	Leu
	380		385		390									
Met	Gly	Leu	Phe	Phe	Gly	Thr	Leu	Thr	Ile	Thr	Leu	Thr	Leu	Gln
	395		400		405									
Asp	His	Ala	Pro	Trp	Val	Pro	Tyr	Leu	Ser	Ile	Val	Gly	Ile	Leu
	410		415		420									
Ala	Ile	Ile	Ala	Ser	Phe	Cys	Ser	Gly	Pro	Gly	Gly	Ile	Pro	Phe
	425		430		435									
Ile	Leu	Thr	Gly	Glu	Phe	Phe	Gln	Gln	Ser	Gln	Arg	Pro	Ala	Ala
	440		445		450									
Phe	Ile	Ile	Ala	Gly	Thr	Val	Asn	Trp	Leu	Ser	Asn	Phe	Ala	Val
	455		460		465									
Gly	Leu	Leu	Phe	Pro	Phe	Ile	Gln	Lys	Ser	Leu	Asp	Thr	Tyr	Cys
	470		475		480									
Phe	Leu	Val	Phe	Ala	Thr	Ile	Cys	Ile	Thr	Gly	Ala	Ile	Tyr	Leu
	485		490		495									
Tyr	Phe	Val	Leu	Pro	Glu	Thr	Lys	Asn	Arg	Thr	Tyr	Ala	Glu	Ile
	500		505		510									
Ser	Gln	Ala	Phe	Ser	Lys	Arg	Asn	Lys	Ala	Tyr	Pro	Pro	Glu	Glu
	515		520		525									

Lys Ile Asp Ser Ala Val Thr Asp Ala Gln Arg Asn  
530 535

<210> 26  
<211> 905  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7611651CD1

<400> 26  
Met Pro Val Arg Arg Gly His Val Ala Pro Gln Asn Thr Tyr Leu  
1 5 10 15  
Asp Thr Ile Ile Arg Lys Phe Glu Gly Gln Ser Arg Lys Phe Leu  
20 25 30  
Ile Ala Asn Ala Gln Met Glu Asn Cys Ala Ile Ile Tyr Cys Asn  
35 40 45  
Asp Gly Phe Cys Glu Leu Phe Gly Tyr Ser Arg Val Glu Val Met  
50 55 60  
Gln Gln Pro Cys Thr Cys Asp Phe Leu Thr Gly Pro Asn Thr Pro  
65 70 75  
Ser Ser Ala Val Ser Arg Leu Ala Gln Ala Leu Leu Gly Ala Glu  
80 85 90  
Glu Cys Lys Val Asp Ile Leu Tyr Tyr Arg Lys Asp Ala Ser Ser  
95 100 105  
Phe Arg Cys Leu Val Asp Val Val Pro Val Lys Asn Glu Asp Gly  
110 115 120  
Ala Val Ile Met Phe Ile Leu Asn Phe Glu Asp Leu Ala Gln Leu  
125 130 135  
Leu Ala Lys Cys Ser Ser Arg Ser Leu Ser Gln Arg Leu Leu Ser  
140 145 150  
Gln Ser Phe Leu Gly Ser Glu Gly Ser His Gly Arg Pro Gly Gly  
155 160 165  
Pro Gly Pro Gly Thr Gly Arg Gly Lys Tyr Arg Thr Ile Ser Gln  
170 175 180  
Ile Pro Gln Phe Thr Leu Asn Phe Val Glu Phe Asn Leu Glu Lys  
185 190 195  
His Arg Ser Ser Ser Thr Thr Glu Ile Glu Ile Ile Ala Pro His  
200 205 210  
Lys Val Val Glu Arg Thr Gln Asn Val Thr Glu Lys Val Thr Gln  
215 220 225  
Val Leu Ser Leu Gly Ala Asp Val Leu Pro Glu Tyr Lys Leu Gln  
230 235 240  
Ala Pro Arg Ile His Arg Trp Thr Ile Leu His Tyr Ser Pro Phe  
245 250 255  
Lys Ala Val Trp Asp Trp Leu Ile Leu Leu Leu Val Ile Tyr Thr  
260 265 270  
Ala Val Phe Thr Pro Tyr Ser Ala Ala Phe Leu Leu Ser Asp Gln  
275 280 285  
Asp Glu Ser Arg Arg Gly Ala Cys Ser Tyr Thr Cys Ser Pro Leu  
290 295 300  
Thr Val Val Asp Leu Ile Val Asp Ile Met Phe Val Val Asp Ile  
305 310 315  
Val Ile Asn Phe Arg Thr Thr Tyr Val Asn Thr Asn Asp Glu Val  
320 325 330  
Val Ser His Pro Arg Arg Ile Ala Val His Tyr Phe Lys Gly Trp  
335 340 345  
Phe Leu Ile Asp Met Val Ala Ala Ile Pro Phe Asp Leu Leu Ile  
350 355 360  
Phe Arg Thr Gly Ser Asp Glu Thr Thr Thr Leu Ile Gly Leu Leu  
365 370 375

Lys	Thr	Ala	Arg	Leu	Leu	Arg	Leu	Val	Arg	Val	Ala	Arg	Lys	Leu
				380					385					390
Asp	Arg	Tyr	Ser	Glu	Tyr	Gly	Ala	Ala	Val	Leu	Phe	Leu	Leu	Met
				395					400					405
Cys	Thr	Phe	Ala	Leu	Ile	Ala	His	Trp	Leu	Ala	Cys	Ile	Cys	Ser
				410					415					420
Leu	Thr	Ser	Val	Gly	Phe	Gly	Asn	Val	Ser	Pro	Asn	Thr	Asn	Ser
				425					430					435
Glu	Lys	Val	Phe	Ser	Ile	Cys	Val	Met	Leu	Ile	Gly	Ser	Leu	Met
				440					445					450
Tyr	Ala	Ser	Ile	Phe	Gly	Asn	Val	Ser	Ala	Ile	Ile	Gln	Arg	Leu
				455					460					465
Tyr	Ser	Gly	Thr	Ala	Arg	Tyr	His	Thr	Gln	Met	Leu	Arg	Val	Lys
				470					475					480
Glu	Phe	Ile	Arg	Phe	His	Gln	Ile	Pro	Asn	Pro	Leu	Arg	Gln	Arg
				485					490					495
Leu	Glu	Glu	Tyr	Phe	Gln	His	Ala	Trp	Ser	Tyr	Thr	Asn	Gly	Ile
				500					505					510
Asp	Met	Asn	Ala	Val	Leu	Lys	Gly	Phe	Pro	Glu	Cys	Leu	Gln	Ala
				515					520					525
Asp	Ile	Cys	Leu	His	Leu	His	Arg	Ala	Leu	Leu	Gln	His	Cys	Pro
				530					535					540
Ala	Phe	Ser	Gly	Ala	Gly	Lys	Gly	Cys	Leu	Arg	Ala	Leu	Ala	Val
				545					550					555
Lys	Phe	Lys	Thr	Thr	His	Ala	Pro	Pro	Gly	Asp	Thr	Leu	Val	His
				560					565					570
Leu	Gly	Asp	Val	Leu	Ser	Thr	Leu	Tyr	Phe	Ile	Ser	Arg	Gly	Ser
				575					580					585
Ile	Glu	Ile	Leu	Arg	Asp	Asp	Val	Val	Val	Ala	Ile	Leu	Gly	Lys
				590					595					600
Asn	Asp	Ile	Phe	Gly	Glu	Pro	Val	Ser	Leu	His	Ala	Gln	Pro	Gly
				605					610					615
Lys	Ser	Ser	Ala	Asp	Val	Arg	Ala	Leu	Thr	Tyr	Cys	Asp	Leu	His
				620					625					630
Lys	Ile	Gln	Arg	Ala	Asp	Leu	Leu	Glu	Val	Leu	Asp	Met	Tyr	Pro
				635					640					645
Ala	Phe	Ala	Glu	Ser	Phe	Trp	Ser	Lys	Leu	Glu	Val	Thr	Phe	Asn
				650					655					660
Leu	Arg	Asp	Ala	Ala	Gly	Gly	Leu	His	Ser	Ser	Pro	Arg	Gln	Ala
				665					670					675
Pro	Gly	Ser	Gln	Asp	His	Gln	Gly	Phe	Phe	Leu	Ser	Asp	Asn	Gln
				680					685					690
Ser	Asp	Ala	Ala	Pro	Pro	Leu	Ser	Ile	Ser	Asp	Ala	Ser	Gly	Leu
				695					700					705
Trp	Pro	Glu	Leu	Leu	Gln	Glu	Met	Pro	Pro	Arg	His	Ser	Pro	Gln
				710					715					720
Ser	Pro	Gln	Glu	Asp	Pro	Asp	Cys	Trp	Pro	Leu	Lys	Leu	Gly	Ser
				725					730					735
Arg	Leu	Glu	Gln	Leu	Gln	Ala	Gln	Met	Asn	Arg	Leu	Glu	Ser	Arg
				740					745					750
Val	Ser	Ser	Asp	Leu	Ser	Arg	Ile	Leu	Gln	Leu	Leu	Gln	Lys	Pro
				755					760					765
Met	Pro	Gln	Gly	His	Ala	Ser	Tyr	Ile	Leu	Glu	Ala	Pro	Ala	Ser
				770					775					780
Asn	Asp	Leu	Ala	Leu	Val	Pro	Ile	Ala	Ser	Glu	Thr	Thr	Ser	Pro
				785					790					795
Gly	Pro	Arg	Leu	Pro	Gln	Gly	Phe	Leu	Pro	Pro	Ala	Gln	Thr	Pro
				800					805					810
Ser	Tyr	Gly	Asp	Leu	Asp	Asp	Cys	Ser	Pro	Lys	His	Arg	Asn	Ser
				815					820					825
Ser	Pro	Arg	Met	Pro	His	Leu	Ala	Val	Ala	Met	Asp	Lys	Thr	Leu
				830					835					840
Ala	Pro	Ser	Ser	Glu	Gln	Glu	Gln	Pro	Glu	Gly	Leu	Trp	Pro	Pro

Leu	Ala	Ser	Pro	845	Leu	His	Pro	Leu	Glu	850	Val	Gln	Gly	Leu	Ile	855	Cys
Gly	Pro	Cys	Phe	860	Ser	Ser	Leu	Pro	Glu	865	His	Leu	Gly	Ser	Val	870	Pro
Lys	Gln	Leu	Asp	875	Phe	Gln	Arg	His	Gly	880	Ser	Asp	Pro	Gly	Phe	885	Ala
Gly	Ser	Trp	Gly	890						895						900	
				905													

&lt;210&gt; 27

&lt;211&gt; 686

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2522075CD1

&lt;400&gt; 27

Met	Ala	Glu	Ala	Ala	Glu	Pro	Glu	Gly	Val	Ala	Pro	Gly	Pro	Gln			
1				5					10					15			
Gly	Pro	Pro	Glu	Val	Pro	Ala	Pro	Leu	Ala	Glu	Arg	Pro	Gly	Glu			
				20					25					30			
Pro	Gly	Ala	Ala	Gly	Gly	Glu	Ala	Glu	Gly	Pro	Glu	Gly	Ser	Glu			
				35					40					45			
Gly	Ala	Glu	Glu	Ala	Pro	Arg	Gly	Ala	Ala	Ala	Val	Lys	Glu	Ala			
				50					55					60			
Gly	Gly	Gly	Gly	Pro	Asp	Arg	Gly	Pro	Glu	Ala	Glu	Ala	Arg	Gly			
				65					70					75			
Thr	Arg	Gly	Ala	His	Gly	Glu	Thr	Glu	Ala	Glu	Glu	Gly	Ala	Pro			
				80					85					90			
Glu	Gly	Ala	Glu	Val	Pro	Gln	Gly	Gly	Glu	Glu	Thr	Ser	Gly	Ala			
				95					100					105			
Gln	Gln	Val	Glu	Gly	Ala	Ser	Pro	Gly	Arg	Gly	Ala	Gln	Gly	Glu			
				110					115					120			
Pro	Arg	Gly	Glu	Ala	Gln	Arg	Glu	Pro	Glu	Asp	Ser	Ala	Ala	Pro			
				125					130					135			
Glu	Arg	Gln	Glu	Glu	Ala	Glu	Gln	Arg	Pro	Glu	Val	Pro	Glu	Gly			
				140					145					150			
Ser	Ala	Ser	Gly	Glu	Ala	Gly	Asp	Ser	Val	Asp	Ala	Glu	Gly	Pro			
				155					160					165			
Leu	Gly	Asp	Asn	Ile	Glu	Ala	Glu	Gly	Pro	Ala	Gly	Asp	Ser	Val			
				170					175					180			
Glu	Ala	Glu	Gly	Arg	Val	Gly	Asp	Ser	Val	Asp	Ala	Glu	Gly	Pro			
				185					190					195			
Ala	Gly	Asp	Ser	Val	Asp	Ala	Glu	Gly	Pro	Leu	Gly	Asp	Asn	Ile			
				200					205					210			
Gln	Ala	Glu	Gly	Pro	Ala	Gly	Asp	Ser	Val	Asp	Ala	Glu	Gly	Arg			
				215					220					225			
Val	Gly	Asp	Ser	Val	Asp	Ala	Glu	Gly	Pro	Ala	Gly	Asp	Ser	Val			
				230					235					240			
Asp	Ala	Glu	Gly	Arg	Val	Gly	Asp	Ser	Val	Glu	Ala	Gly	Asp	Pro			
				245					250					255			
Ala	Gly	Asp	Gly	Val	Glu	Ala	Gly	Val	Pro	Ala	Gly	Asp	Ser	Val			
				260					265					270			
Glu	Ala	Glu	Gly	Pro	Ala	Gly	Asp	Ser	Met	Asp	Ala	Glu	Gly	Pro			
				275					280					285			
Ala	Gly	Arg	Ala	Arg	Arg	Val	Ser	Gly	Glu	Pro	Gln	Gln	Ser	Gly			
				290					295					300			
Asp	Gly	Ser	Leu	Ser	Pro	Gln	Ala	Glu	Ala	Ile	Glu	Val	Ala	Ala			
				305					310					315			
Gly	Glu	Ser	Ala	Gly	Arg	Ser	Pro	Gly	Glu	Leu	Ala	Trp	Asp	Ala			



	320		325		330
Ala Glu Glu Ala	Glu Val Pro Gly Val	Lys Gly Ser Glu Glu	Ala		
	335		340		345
Ala Pro Gly Asp	Ala Arg Ala Asp Ala	Gly Glu Asp Arg Val	Gly		
	350		355		360
Asp Gly Pro Gln	Gln Glu Pro Gly Glu	Asp Glu Glu Arg Arg	Glu		
	365		370		375
Arg Ser Pro Glu	Gly Pro Arg Glu Glu	Glu Ala Ala Gly Gly	Glu		
	380		385		390
Glu Glu Ser Pro	Asp Ser Ser Pro His	Gly Glu Ala Ser Arg	Gly		
	395		400		405
Ala Ala Glu Pro	Glu Ala Gln Leu Ser	Asn His Leu Ala Glu	Glu		
	410		415		420
Gly Pro Ala Glu	Gly Ser Gly Glu Ala	Ala Arg Val Asn Gly	Arg		
	425		430		435
Arg Glu Asp Gly	Glu Ala Ser Glu Pro	Arg Ala Leu Gly Gln	Glu		
	440		445		450
His Asp Ile Thr	Leu Phe Val Lys Ala	Gly Tyr Asp Gly Glu	Ser		
	455		460		465
Ile Gly Asn Cys	Pro Phe Ser Gln Arg	Leu Phe Met Ile Leu	Trp		
	470		475		480
Leu Lys Gly Val	Ile Phe Asn Val Thr	Thr Val Asp Leu Lys	Arg		
	485		490		495
Lys Pro Ala Asp	Leu Gln Asn Leu Ala	Pro Gly Thr Asn Pro	Pro		
	500		505		510
Phe Met Thr Phe	Asp Gly Glu Val Lys	Thr Asp Val Asn Lys	Ile		
	515		520		525
Glu Glu Phe Leu	Glu Glu Lys Leu Ala	Pro Pro Arg Tyr Pro	Lys		
	530		535		540
Leu Gly Thr Gln	His Pro Glu Ser Asn	Ser Ala Gly Asn Asp	Val		
	545		550		555
Phe Ala Lys Phe	Ser Ala Phe Ile Lys	Asn Thr Lys Lys Asp	Ala		
	560		565		570
Asn Glu Ile His	Glu Lys Asn Leu Leu	Lys Ala Leu Arg Lys	Leu		
	575		580		585
Asp Asn Tyr Leu	Asn Ser Pro Leu Pro	Asp Glu Ile Asp Ala	Tyr		
	590		595		600
Ser Thr Glu Asp	Val Thr Val Ser Gly	Arg Lys Phe Leu Gly	Gly		
	605		610		615
Asp Glu Leu Thr	Leu Ala Asp Cys Asn	Leu Leu Pro Lys Leu	His		
	620		625		630
Ile Ile Lys Ile	Val Ala Lys Lys Tyr	Arg Asp Phe Glu Phe	Pro		
	635		640		645
Ser Glu Met Thr	Gly Ile Trp Arg Tyr	Leu Asn Asn Ala Tyr	Ala		
	650		655		660
Arg Asp Glu Phe	Thr Asn Thr Cys Pro	Ala Asp Gln Glu Ile	Glu		
	665		670		675
His Ala Tyr Ser	Asp Val Ala Lys Arg	Met Lys			
	680		685		

&lt;210&gt; 28

&lt;211&gt; 2984

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7475353CB1

&lt;400&gt; 28

gttggcagaa ggggtcccggg cccagagcca gcggggccgt gctgagacgg cgtacgtgcc 60  
ctgcgtgagt gcgtggcggc ggcgcgtgcg ctaggggagt gggcggtgag gcctggtcca 120  
cgtgcgtccc ttcccgggac ccccgagct tggcgcccag cggctacgtg agccaaggca 180

```

ccccgatgtc cgcgccccctc tccgagtgac aagtccccggc ctccgggtccc gcagtgcccc 240
cagcctcggc cggcgctccac gcattgccat ggtgactgtg ggcaactact gcgaggccga 300
agggcccggtg ggtccggcct ggatgcagga tggcctgagt cctgcttct tcttcacgct 360
cgtgccctcg acgcgggatgg ctctagggac tctggccttg gtgctggctc ttccctgcag 420
acgccgggag cggcccgctg gtgctgattc gctgtcttgg ggggcccggcc ctgcgcatctc 480
tccctacgtg ctgcagctgc ttctggccac acttcaggcg gcgctgcccc tggccggcct 540
ggctggccgg gtgggcaactg cccgggggggc cccactgcca agctatctac ttctggcctc 600
cgtgctggag agtctggccg ggcctgtggg cctgtggctg cttgtcgtgg agcggagcca 660
ggcacggcag cgtctggcaa tgggcatctg gatcaagttc aggcacagcc ctggtctctc 720
gctcctctgg actgtggcgt ttgcagctga gaacttggcc ctggtgtctt ggaacagccc 780
acagtgggtg tgggcaaggg cagacttggg ccagcaggtt cagtttagcc tgggggtgct 840
gcggtatgtg gtctctggag ggctgtttgt cctgggtctc tgggccccctg gacttcgtcc 900
ccagtcctat acattgcagg ttcattgaaga ggaccaagat gtggaaagga gccaggttcg 960
gtcagcagcc caacagtcta cctggcgaga ttttggcagg aagctccgcc tccctgagtgg 1020
ctacctgtgg cctcgaggga gtccagctct gcagctgggt gtgctcatct gcctggggct 1080
catgggtttg gaacgggcac tcaatgtgtt ggtgcctata ttctatagga acattgtgaa 1140
cttgctgact gagaaggcac cttggaactc tctggcctgg actgttacca gttacgtctt 1200
cctcaagttc ctccaggggg gtggcactgg cagtacaggc ttctgtagca acctgcgcac 1260
cttcctgtgg atccgggtgc agcagttcac gtctcggcgg gtggagctgc tcactctctc 1320
ccacctgcac gagctctcac tgcgctggca cctggggcgc cgcacagggg aggtgctgcg 1380
gatcgcgag cggggcacat ccagtgtcac agggctgctc agctacctgg tgttcaatgt 1440
catccccacg ctggccgaca tcatcattgg catcatctac ttcagcatgt tcttcaacgc 1500
ctggtttggc ctcatgtgtt tctgtgtgat gagtctttac ctaccctga ccattgtggt 1560
cactgagtgg agaaccaagt ttctgtgtgc tatgaacaca caggagaacg ctaccggggc 1620
acgagcagtg gactctctgc taaacttcga gacggtgaag tattacaacg ccgagagtta 1680
cgaagtggaa cgctatcgag aggccatcat caaatatcag ggtttggagt ggaagtcgag 1740
cgcttccactg gttttactaa atcagaccca gaacctgggt attgggctcg ggctcctcgc 1800
cggctccctg ctttgcgcat actttgtcac tgagcagaag ctacaggttg gggactatgt 1860
gctctttggc acctacatta tccagctgta catgccccct aactggtttg gcacctacta 1920
caggatgata cagaccaact tcattgacat ggagaacatg tttgacttgc tgaagagga 1980
gacagaagtg aaggaccttc ctggagcagg gcccttctgc tttcagaagg gccgtattga 2040
gtttgagaac gtgcacttca gctatgccga tgggcccggg actctgcagg acgtgtcttt 2100
cactgtgatg cctggacaga cacttgcctt ggtgggcccc tctggggcag ggaagagcac 2160
aattttgcgc ctgctgtttc gcttctacga catcagctct ggctgcatcc gaatagatgg 2220
gcaggacatt tcacagggtg cccaggcctc tctccggtct cacattggag ttgtgcccc 2280
agacactgtc ctctttaatg acaccatcgc cgacaatatc cgttacggcc gtgtcacagc 2340
tgggaatgat gaggtggagg ctgctgtcca ggctgcaggc atccatgatg ccattatggc 2400
tttccctgaa ggggtacagga cacagggtgg cgagcgggga ctgaagctga gcggcgggga 2460
gaagcagcgc gtcgccattg cccgcacctt cctcaagggt cccggcatca ttctgctgga 2520
tgaggcaacg tcagcgctga atacatctaa tgagagggoc atccaggctt ctctggccaa 2580
agtctgtgcc aaccgcacca ccatcgtagt ggcacacagg ctctcaactg tggtaactgt 2640
tgaccagatc ctctgcatca aggatggctg catcgtggag aggggacgac acgaggctct 2700
gttgctccga ggtggggtgt atgctgacat gtggcagctg cagcagggac aggaagaaac 2760
ctctgaagac actaagcctc agaccatgga acggtgacaa aagtttggcc acttccctct 2820
caaagactaa cccagaaggg aataagatgt gtctccttct cctggcttat ttcactctgg 2880
tcttggggta tgggtgctagc tatggttaagg gaaagggacc tttccgaaaa acatcttttg 2940
gggaaataaa aatgtggact gtgaaaaaaa aaaaaaaaaa aaaa 2984

```

&lt;210&gt; 29

&lt;211&gt; 1846

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3107278CB1

&lt;400&gt; 29

```

aatactatca gtcttccctg cgtacggctg gaactattcc ccttcgccac tgccectgg 60
gaggctgcgg gcaacggagc aacagcagcg gcgcgagcg aggcgaacac caccctcat 120
cccctccgga caagggggac aacgcctcca actgtgactg ccgcgcatgg gactacggca 180
tccgcgccgg cctcgtccag aacgtggtca gcaagtggga tctgtgtgtg ataatgcctg 240
gaagggtccat atcgctaagt tctccttact ggtggattaa tctttggtac ctaataactg 300

```

gatgcattgc	tgactgggtc	ggccggcggc	ctgtgctgct	gttttccatc	atcttcatte	360
tgatcttttg	actgactgtg	gcactgtcag	tgaatgtgac	aatgttcagc	acactcaggt	420
tctttgaagg	attttgcttg	gctggaatca	ttctcacctt	gtatgcttta	cgaatagagc	480
tgtgcccccc	tggaaaacgg	ttcatgatta	cgatggtggc	gagcttcgtg	gccatggcgg	540
gccagttcct	catgcctggg	ctagccggcc	tgtgccggga	ttggcaggtg	ctgcaggccc	600
tcacatctctg	ccccttctctg	ctcatgctgc	tctactggtc	gatattcccc	gagtcctctc	660
ggtggctaata	ggccacccag	cagtttgagt	ctgcaaagag	gctgatcctc	cacttcacac	720
agaagaatcg	catgaaccct	gagggcgaca	tcaaggggtg	gataccagag	ctggagaaag	780
agctttcccg	gagggccaag	aaggtctgca	togtgaaggt	ggtggggaca	cggaacctgt	840
ggaagaacat	tgtggtcctg	tgtgtgaact	cgctgacggg	gtacgggatc	caccactgct	900
ttgccaggag	catgattggc	cacgaggtga	aggtgccgct	cctggagaac	ttctatgctg	960
actactatac	cacggccagc	atcgcgctgg	tgtcctgcct	ggccatgtgc	gtggtggtcc	1020
gattcctcgg	gcgcaggggg	gggtgctgct	tcttcatgat	cctcaccgcc	ctggcctcgc	1080
tcctgcagct	cggcctcctc	aacctgattg	gaaagtacag	ccagcaccca	gactcaggga	1140
tgagtgcagc	cgtcaaggac	aaattttcca	tgcggttttc	catcgtgggc	atgtttgctt	1200
cccatgcggt	ggggagcctc	agcgtgttct	tctgtgcgga	gatcaccccg	acgggtgata	1260
ggtgtggcgg	gctggggctg	gtgctggcca	gcgcgggctt	cggcatgctg	acggcaccca	1320
tcacgcagct	gcacaaccag	aaaggctact	tccctgcacca	catcatcttt	gctgtctgca	1380
cgctcatctg	catcatctgc	atcctcctgc	tgcgcgagag	cagggaccag	aacctgcctg	1440
agaacatttc	taacggggag	cactacacgc	gccagccgct	gctgccgcac	aagaaggggg	1500
agcagccact	gctgctcacc	aacgcccagc	tcaaggacta	ctcgggcctc	cacgatgccg	1560
cagccgcggg	tgacacactg	cccgagggtg	ccacggccaa	cggcatgaag	gccatgtagc	1620
ccggcctcgc	gaaccggggg	ctccagggtc	tggggcagct	tgggcacagg	tttacagacc	1680
aggggacgaa	cacgcagcca	gggggtggaa	agatgacatc	agccaagctg	agcctctcaa	1740
ctggtgtggg	gaaatcctgt	ctttccaaaa	gtccaaggag	cgcggttcgg	aggagacaaa	1800
ctctttggaa	ataacccttt	caagactttc	ttttctgccg	ttaaaa		1846

&lt;210&gt; 30

&lt;211&gt; 1458

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7473394CB1

&lt;400&gt; 30

atgcagaata	ttaccaaaga	atttgggaaca	ttcaaggcaa	atgacaacat	caattttacaa	60
gtaaaggcag	gagagattca	tgcgttgctt	ggagaaaacg	gtgctggcaa	atctacattg	120
atgaacgtgc	tttccggatt	attagagccg	acatcaggga	aaattttgat	gcgtgggaaa	180
gaagtacaga	tcacaagccc	gacaaaagcc	aatcaattag	ggattgggat	ggtccatcag	240
cactttatgc	ttgttgcagc	ctttactgtc	acagaaaaca	tcgtgttggg	aagcgaacct	300
agtcgtgcag	ggatgcttga	ccataaaaaa	gcgcgaaaag	agatccaaaa	agtttctgaa	360
caatatggat	tatcagtcaa	cccggatgct	tatgttcgtg	atatttcagt	tggtgatggaa	420
caacgggtag	aaatttttaa	aacactttac	cgaggagcag	atgtactgat	ttttgatgag	480
ccgacagctg	tattgacccc	tcaggaaatt	gatgaattaa	tcgtgatcat	gaaggaatta	540
gtcaaagaag	gcaagtcaat	catttttgatt	acgcataagt	tagatgaaat	caaagcagta	600
gctgaccgtt	gtacagttaa	ccgccgtgga	aaaggaatcg	gtacagtcaa	cgttaaagac	660
gttacctcac	agcaattagc	tgatatgatg	gtcgggaagag	cggtttcatt	caaaacgatg	720
aaaaaagaag	cgaagcctca	agaagtcgtt	ttgtctattg	aaaatctagt	ggtaaaagaa	780
aatcgtggat	tagaagccgt	gaaaaacctg	aacttagagg	ttcgtgctgg	cgaagtactt	840
ggtatcgctg	gaatcgatgg	aaacgggcag	tcggagttga	tccaagcttt	gactggtttg	900
cgaaaggcag	aaagcggaca	tatcaagcta	aaaggggaag	acatcaccaa	taaaaaacct	960
cgaaagatca	ctgaacatgg	tgtaggacat	gtgccagaag	accgtcataa	atacgggttg	1020
gtcctagata	tgaaaacatt	gccctgcaaa	cgtatcatca	aaaaccttac	1080	
agtaaaaacg	gtatgctgaa	ttattcagtg	ataaatgaac	atgccagaga	attgatcgaa	1140
gaatatgatg	ttcgaacaac	gaatgaactt	gttcctgcaa	aagctttatc	aggcggaaat	1200
cagcaaaaag	caatcatcgc	tcggatagtc	gaccgagatc	ctgatctgtt	gatcgttgca	1260
aatccaactc	gtgggctgga	tgtaggtgcg	atcgaattta	ttcataaacg	tctgatcgaa	1320
caaagggaca	aatacaaagc	agtgttattg	attagtttcg	aattagaaga	aatttttaaa	1380
gttttcggatc	gtattgctgt	tatccatgaa	ggagaaatcg	tcgggatcgt	tgatccgaaa	1440
gaaacatctg	aaaattaa					1458

<210> 31  
<211> 1234  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7473900CB1

<400> 31  
atgaagtcgg gtcctggcat ccaagccggc atcgacctca cagcggggggc cgcagggggg 60  
acagcgtgtg tactgactgg gcagcccttc gacacaataa aagtgaagat gcagacgttc 120  
cctgacctgt acaagggcct caccgactgc ttcctgaaga catagcccca agtgggtctc 180  
cggggcttct acaagggcac cggcccggca cttatggcct acgtcgccga aaactcggtc 240  
ctcttcattgt gctacgggtt ctgccagcag tttgtcagga aagtggctgg aatggacaag 300  
caggcaaaagc tgagtgatct ccagactgca gcccggggt ccttcgcctc tgcatttgct 360  
gcactggctc tctgccccac tgagcttggt aagtgcgggc tacagaccat gtatgaaatg 420  
gagatgtcag ggaagatagc aaaaagccat aatacaattt ggtctgtcgt gaagggtatc 480  
cttaaaaagg atggccctt gggcttctac catggactct cgagtactct acttcaagaa 540  
gtaccgggtt atttcttttt ctttgggtggc tatgaactga gccgatcgtt ttttgctgca 600  
gggagatcaa aagatgaact aggcctgtgc catttgatgt taagtgggtg agttgctgga 660  
atttgctgtg ggtctgtcgt gttcccagtg gattgtatta aatccagaat tcaagttctt 720  
tccatgtatg ggaaacaggc aggatttatt ggtaccctct taagtgttgt gagaaatgaa 780  
ggaatagtag ccttatattc tggactgaaa gctactatga ttcgagcaat ccctgccaat 840  
ggggcactgt ttgtggccta cgaatacagc aggaagatga tgatgaaaca gttggaagca 900  
tactgaagtg tcttggtgaa cctggatccg agtccatgag tttgaggact acagttcatc 960  
acagggttca gcagagtaca agaccactgt ctaattttga cttcatggga attttgggtt 1020  
tatcttccct tcttctaccc taaatcttaa ctttatggaa gggcctctat tttacatcat 1080  
ataatttctg ccataaattg tattgaaata ggaaagtgc tgctcttgca cttgctggaa 1140  
tgtacagggg gggctggttg gccctatgta cctaactcga aaaactaaat atcgttctgt 1200  
cagggccttt gcataaagcc atttgtgtgt acat 1234

<210> 32  
<211> 1255  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7475045CB1

<400> 32  
gtgacctttc cccccagatc ccagggtccag gccgcgcctc ggctggcagg tgtgggcaca 60  
gaggcagctg ggattggtcg cagctggcgg aggcgcgtcc caggctccgg cagaccgctg 120  
gaacagttga gccagagcag gtggactgct gagatagacc agggacacca ggcagccaca 180  
ggcctgtcag accaggaccc ttacctcta gacatggcct cggctccctg caaaccacag 240  
ccccgtagcc ctgcgaggtt acagacagcc taaacgccac caccacaggg cctgtgccgt 300  
gccctgacc cgggcacaga aggcactgg cccggaggcc atggagacgg tgccccagc 360  
agtggacctg gtgtggtg ctctgtcctg ctgcctggcc tgtgtcttca ccaacccct 420  
ggaggtggtg aagacgcggc tgcagctgca gggggagctg caggcccggg gcacctacc 480  
acggccctac catggcttca tagcctctgt cgctgctgtg gcccgagcag acgggctgtg 540  
gggcctgcag aaggggctgg ctgcccgcct tctgtaccaa ggcctcatga atggcgctcg 600  
tttctactgc tacagcctgg cgtgccaggg ttgcctcacg cagcaaccag gtggcaccgt 660  
ggttgccgga gccgtggcgg gggcactggg agccttcgtg gggagccctg cttacctgat 720  
caaacgcag ctgcaagctc agacagtggc cgcagtggcc gtgggacacc agcacaatca 780  
ccagactgtc ctgggtgcct tggagaccat ctggcggcag caagggctct tggggctgtg 840  
gcagggcggtt ggtggggctg tgccccagat catggtgggc tcagctgccc agctggccac 900  
cttcgcctct gccaaaggct gggtagagaa gcaacagtgg ctccctgagg acagctggct 960  
ggtggccctg gctgggggca tgatcagcag catagccgtg gttgtcgtca tgactccct 1020  
cgatgtgggtc agcacgcggc tatacaatca gccggtggac acagctggca ggggccagct 1080  
ctatgggggc ctaccgact gcatggtgaa gatctggcgg caggagggcc ccctggcact 1140  
ctacaagggc ctgggccccg cctacctgcg cctgggcccc cacaccatcc tcagcatgct 1200  
cttctggggc gagcttcgga aactggctgg gcgggcccag cacaagggca cctag 1255

<210> 33  
<211> 957  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7475611CB1

<220>  
<221> unsure  
<222> 1  
<223> a, t, c, g, or other

<400> 33  
ngccgcgctg gccaccttgc ccataaaacg caccggttcg gtgcgataca agatcatcgt 60  
cgatcatcgt atcgctgtcc tgtgggtgat cagttggacg acgaccggaa ggattttcag 120  
atgagtgcca aggtcctgct gtcgaccgag cacctgtacg ccacccaccc gggccgctct 180  
atgggtactga ccgacgttaa tgtctccttt cgcgcggggg ttgcgctagc gatcctggga 240  
gctaattggat ccggttaagac gacctcatg cgtgcctgt ccggttcctt caaacccgcc 300  
aagggtcacg tcaagagggg cgacatcgtt gtcagctacg ggcgcgctca acttcgtgag 360  
caccgtcgag ccgtccagct tgtgttgcaa gacctgacg accagctctt tagcgccgat 420  
gtcagccagg atgtctcctt cggccccatg aatatggggc tcaaagttag cgaggtgctg 480  
gaccgggtct ccgagtcctt agaactgctc gggggccagc atctggctga gcgtgccacg 540  
tatcaactgt cctatgggtga gcgcaagagg gtcgcggttg ccggtgccgt ggccatgcgc 600  
ccggtactgc tgcctcttga tgagcccacc gccggacttg acccggttgg agtcacccag 660  
atgttggagg ccctggatcg gctgcgcgat catggaacaa cgggtggcgt ggctacccac 720  
gacgtcgacc tggctctggc gtggggcgag gaggcccttg tcgttgctga cggtcagggtg 780  
caccaaggac cgatcggcga gttacttgcc gatgccgaca ccgtgggacg ggcacacctg 840  
caccttcctg ggcccccg gctcgcccg cgctcggtg ttccgggacct tcccaggacg 900  
atggacgacg tcgtggcgat gctgtccgac aatccctcgc cagctccctc gaattga 957

<210> 34  
<211> 2407  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7475617CB1

<400> 34  
gcggcgccgg cctcggcctc ctctctctggg gcggcgccgg aggacagcag cgccatggag 60  
gagctcgcta ctgagaagga ggcggaggag agccaccggc aagacagcgt gagcctgctc 120  
accttcatcc tgctgtcac gctcaccatc ctaccatct ggctcttcaa gcaccgccgg 180  
gtgcgctttc tgcacgagac cgggctggcc atgatctatg ggctcatcgt tggggtgatc 240  
ctgagggtatg gtacccctgc taccagtggc cgtgacaaat cactcagctg cactcaggaa 300  
gacagggcct tcagtacctt attagtgaat gtcagcggaa agttcttcga atacactctg 360  
aaaggagaaa tcagtcctgg caagatcaac agcgtagagc agaatgatat gctacggaag 420  
gtaacattcg atccagaagt atttttcaac attctctctg ctccaattat ttttcatgct 480  
ggatacagct taaagaagag acactttttc agaaatcttg gatctatact ggcctatgcc 540  
ttcttgggga ctgctgtttc atgcttcatt attggaaatc tcatgtatgg tgtggtgaag 600  
ctcatgaaga ttatgggaca gctctcagat aaattttact acacagattg tctctttttt 660  
ggagcaatca tctctgccac tgacccagtg actgtgctgg cgatatttaa tgaattgcat 720  
gcagacgtgg atctttacgc acttcttttt ggagagagcg tccataatga tgctgttgcc 780  
attgtactgt cctcgtctat tgttgccctac gaccagcggg gactgaacac tcacgccttt 840  
gatgctgctg ccttttttaa gtcagttggc atttttctag gtatatattg tggctctttt 900  
accatgggag ctgtgactgg tgttgtgact gctctagtga ctaagtttac caaactgcac 960  
tgcttcccc tgcgtggagac ggcgctgttc ttctcatgt cctggagcac gtttctcttg 1020  
gcagaagcct gcggatttac aggtgttgta gctgtccttt tctgtggaat cacacaagct 1080  
cattacacct acaacaatct gtcggtggaa tcaagaagtc gaaccaagca cctctttgag 1140  
gtgttacatt tcctggcaga gaacttcata ttctcttaca tgggcctggc actgtttacc 1200  
ttccagaagc acgttttcag ccccattttc atcatcggag cttttgttgc catcttctctg 1260

```

ggcagagccg cgcacatcta cccgctctcc ttcttcctca acttgggcag aaggcataag 1320
attggctgga attttcaaca catgatgatg ttttcaggcc tcaggggagc aatggcattt 1380
gcgttggcca tccgtgacac ggcatectat gctcgccaga tgatgttcac gaccaccctt 1440
ctcatttgtt tcttcactgt ctggatcatt ggaggaggca cgacacccat gttgtcatgg 1500
cttaacatca gagttggcgt cgaggagccc tccgaagagg accagaatga acaccactgg 1560
cagtacttca gagttggtgt tgaccccgat caagaccac caccacaaca cgacagcttt 1620
caagtcttac aaggggacgg cccagattct gccagaggaa accggacaaa acaggagagc 1680
gcatggatat tcaggctgtg gtacagcttt gatcacaatt atctgaagcc catcctcaca 1740
cacagtggtc cccactaac caccacgctc ccgcctggt gtggcttact agctcgatgt 1800
ctgaccagtc cccaggtgta cgataaccaa gagccactga gagaggaga ctctgatttc 1860
atcctgaccg aaggcgacct gacattgacc tacgggggaca gcacagtgc tgcaaatggc 1920
tcctcaagtt cgcacaccgc ctccacgagt ctggagggca gccggagAAC gaagagcagc 1980
tcggagggaag tgctggagcg agacctggga atgggagacc agaaggtttc gagccggggc 2040
acccgcctag tgtttccctt ggaagataat gcttgacttt ccccccaagc cctggcgcgga 2100
tggggtaggg tcccgatggg actgaagatt tgaaaataca tccacgaaca tttcaacatg 2160
gaacgaagaa ttatagtccc ttctctggct atatccataa agaacytagt tatgaaatgt 2220
ttaaaccaca aggcacaaatg tgttatactc ttattttttg attaatctga ggaaaggagg 2280
tatttagaaa ctgatgatgg tatcactgca aaaagcattc aacttttttt tttttggtgg 2340
agatggtgtt ccactctgtc acccaggctg gagtgcggtg ccttggtctg ggcccactgc 2400
aacctct 2407

```

&lt;210&gt; 35

&lt;211&gt; 2767

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7473314CB1

&lt;400&gt; 35

```

atggagggct ctggggggcg tgccggggcag cgggcgcgcg tgctggggcg gggcggggag 60
gcggcgggcg cgccggctgg ggccgttcgc ggccggcgcg cggcggtgcg ggccgtgctg 120
ctgacggagc tgctggagcg cgcgcgtttc tacggcatca cgtccaacct ggtgctattc 180
ctgaacgggg cgccgttctg ctggggaggc gcgcaggcca gcgaggcgct gctgctcttc 240
atgggcctca cctacctggg ctgcgcgttc ggaggctggc tggccgacgc gggctggggc 300
cgggcgcgcg ccacctctgt gagcctggcg ctctacctgc tgggcatgct ggccttcccg 360
ctgctggccg cgcccgccac gcgagccgcg ctctgcggtt ccgcgcgcct gctcaactgc 420
acggcgccctg gtcccgaacg cgccgcgcgc tgctgctcac cggccacctt cgccgggctg 480
tgctggtggg gcttggcggt ggccacgcgc aaggccaaca tcacgccctt cgccgcgcag 540
caggttaaag atcggaggtc ggaagccact aggagatttt ttaattggtt ttattggagc 600
attaacctgg gagcgatcct gtcgttaggt ggcattgcct atattcagca gaacgtcagc 660
tttgtcactg gttatgcgat cccactgtc tgcgtcggcc ttgcttttgt ggccttcttc 720
tgtggccaga gcgttttcat caccaagcct cctgatggca gtgccttcac cgatagtgtc 780
aagatactga cgtattcctg ctgttcccag aagcgaagtg gagagcgcca gagtaatggt 840
gaaggcattg gagtctttca gcaatcttct aaacaaagtc tgtttgattc atgtaagatg 900
tctcatgggt ggccattttac agaagagaaa gtggaagatg tgaaagctct ggtcaagatt 960
gtccctgttt tcttggcctt gataccttac tggacagtgt atttccaaat gcagacaaca 1020
tatgttttac agagtcttca tttgaggatt ccagaaattt caaatattac aaccactcct 1080
cacacgctcc ctgcagcctg gctgaccatg tttgatgctg tgctcatcct cctgctcatc 1140
cctctgaagg acaaactggt cgatcccatg ttgagaagac atggcctgct cccatcctcc 1200
ctgaagagga tcgcctgggg catgttcttt gtcattgtgt cagcctttgc tgcagggaatt 1260
ttggagagta aaaggctgaa ccttgtttaa gagaaaacca ttaatcagac catcggaac 1320
gtcgtctacc atgctgccga tctgtcgtg ttgtggcagg tgccgcagta cttgctgatt 1380
gggatcagcg agatctttgc aagtatcgca ggccctggaat ttgcatactc agctgcccc 1440
aagtccatgc agagtgcct aatgggcttg ttctttttct tctctggcgt cgggtcgttc 1500
gtgggttctg gactgctggc actggtgtct atcaaagcca tcggatggat gagcagtcac 1560
acagactttg gtaatatata cggctgctat ttgaactatt actttttcct tctggctgct 1620
attcaaggag ctacctcct gcttttctc attatttctg tgaaatatga ccatcatcga 1680
gaccatcagc gatcaagagc caatggcgtg cccaccagca ggagggcctg acctcctga 1740
ggccatgtgc ggtttctgag gctgacatgt cagtaactga ctggggtgca ctgagaacag 1800
gcaagacttt aaattcccat aaaatgtctg acttactga aacttgcctg ttgctggat 1860
tgatttcttc tttccctcta tccaaaggag cttggttaagt gccttactgc agcgtgtctc 1920

```

```

ctggcacgct gggccctccg ggaggagagc tgcagatttc gagtatgtcg cttgtcattc 1980
aaggtctctg tgaatcctct agctgggttc cctttttttac agaaactcac aaatggagat 2040
tgcaaagtct tggggaactc cacgtgttag ttggcatccc agtttcttaa acaaatagta 2100
tcacctgctt cccatagcca tatctcactg taaaaaaaaa aattaataaa ctgttactta 2160
tatttaagaa agtgaggatt tttttttttt aaagataaaa gcatggtcag atgctgcaag 2220
gattttacat aaatgccata tttatggttt ccttccctgag aacaatcatg ctcttgccat 2280
gttctttgat ttaggctggg agtaaacaca tttcatctgc tgcttcaaaa agtacttact 2340
ttttaaacca tcaacattac ttttctttct taaggcaagg catgcataag agtcatttga 2400
gaccatgtgt cccatctcaa gccacagagc aactcacggg gtacttcaca ccttacctag 2460
tcagagtgtt tatatatagc tttatttttg tacgattgag actaaagact gatcatgggt 2520
gtatgtaagg aaaacattct tttgaacaga aatagtgtaa ttaaaaaataa ttgaaagtgt 2580
taaagtgtgaa cttagagctgt ttgaccagtc acatttttgt attgttactg tacgtgtatc 2640
tggggcttct cgttttggtt atactttttc tgtatttgtt gctgtatttt tggcataact 2700
ttattataaa aagcatctca aatgcgaaat ccaaaaaaaa aaaaaaaaaa gatcggccgc 2760
aagctta 2767

```

&lt;210&gt; 36

&lt;211&gt; 2182

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 70356714CB1

&lt;400&gt; 36

```

gcagtcgcgt cagccgagggc agctctgttc atggcggttct cgaagctctt ggagcaagcc 60
ggaggcggtg gcctcttcca gacctgcag gtgctcacct tcatcctccc ctgcctcatg 120
ataccttccc agatgtctct ggagaacttc tcagccgcca tcccaggcca ccgatgctgg 180
acacacatgc tggacaatgg ctctgcggtt tccacaaaca tgacccccaa ggcccttctg 240
accatctcca tcccgcagg cccaaccag gggccccacc agtgccgccc ctctcgccag 300
ccacagtggc agctcttggg cccaatgcc acggccacca gctggagcga agctgacacg 360
gagccgtgtg tggacggctg ggtctatgac cgcagcgtct tcacctccac catcgtggcc 420
aagtgggacc tgggtgtgcag ctcccagggc ttgaagcccc taagccagtc catcttcatg 480
tccgggatcc tgggtgggctc ctttatctgg ggcctcctct cctaccgggt tgggaggaag 540
ccgatgctga gctgggtgct cctgcagttg gccgtggcgg gcaccagcac catcttcgcc 600
ccaacattcg tcatctactg cggcctgctg ttctgtggccg cttttgggat ggccggcatc 660
ttctgagtt cactgacact gatggtggag tggaccacga ccagcaggag ggcggtcacc 720
atgacgggtg tgggatgtgc cttcagcgca ggccaggcgg cgctgggccc cctggccctt 780
gccctgcggg actggaggac tctccagctg gcagcatcag tgcccttctt tgccatctcc 840
ctgatatact ggtggctgcc agaatccgcc cgggtgctga ttattaaggg caaaccagac 900
caagcacttc aggagctcag aaaggtggcc aggataaatg gccacaagga ggccaagaac 960
ctgaccatag aggtgctgat gtccagcgtg aaaggaggag tggcctctgc aaaggagccg 1020
cggtcggtgc tggacctgtt ctgctgccc gtgctccgct ggaggagctg cgccatgctg 1080
gtgggtgaatt tctctctatt gatctctac tatgggctgg tcttcgacct gcagagcctg 1140
ggcgtgaca tcttctctct ccaggccctc ttccggggccg tggacttctt gggccggggc 1200
accactgccc tcttgtctcag tttccttggc cgccgcacca tccaggcggg ttcccaggcc 1260
atggccggcc tcgccattct agccaacatg ctggtgccgc aagatttgca gaccctgcgt 1320
gtggtctttg ctgtgctggg aaagggatgt tttgggataa gcctaacctg cctcaccatc 1380
tacaaggctg aactctttcc aacgccagt cggatgacag cagatggcat tctgcataca 1440
tgaggccggc tgggggctat gatgggtccc ctgatcctga tgagccgcca agccctgccc 1500
ctgctgcctc ctctctctta tggcgttatc tccattgctt ccagcctggt tgtgctgttc 1560
ttctcccgag agaccaggg acttccgctc cctgacacta tccaggacct ggagagccag 1620
aaatcaacag cagcccaggg caaccggcaa gaggccgtca ctgtggaaag tacctcgctc 1680
tagaaattgt gcctgcatgg agccccctta gtcaaagact cctggaaagg agttgcctct 1740
tctccaatca gagcgtggag ggcgagtg ggcctggcat ggcaaggcc 1800
agggcagcgt ggcgagtg acagcgtggc cgtctgctgt ggctgaaggc agcttcacca 1860
gctcactcct cttctccctg ccctgatcag attccccacc ttaccggggc cctacaggag 1920
cctgtgcaga tggccatgcc caaccaataa cgagacgggt cccctccctt tccctgccag 1980
getcatgtct ttacaccttc actcagccac gccaaaccaga gactgggttc caatctcacc 2040
ccaccacata cagagccctc atctgtgaaa tgagaatgat cacgtgacct acccccagag 2100
gcaggtatca ggggtgaactg atcttagcac cggccaaata aatggaacct gctgagagag 2160
ctgccagaaa aaaaaaaaaa aa 2182

```

<210> 37  
<211> 2811  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7611491CB1

<400> 37

```
cagacggggc tggggcaggc atggcggatt ccagcgaagg cccccgcgcg gggcccgggg 60
aggtggctga gctccccggg gatgagagtg gcaccccagg tggggaggct tttcctctct 120
cctccctggc caatctgttt gagggggagg atggctccct ttcgccctca ccggtgatg 180
ccagtcgccc tgctggccca ggcgatgggc gaccaaactc gcgcatgaag ttccaggcgc 240
cttcgcgaag ggggtgccc aacccatcga tctgctggag tccaccctat atgagtcctc 300
ggtggtgcct gggcccaaga aagcacccat ggactcactg tttgactacg agccgcagag 360
tcaccactcc agtgacaaca agaggtggag gaagaagatc atagagaagc agccgcagag 420
ccccaaagcc cctgcccctc agccgcccc catcctcaaa gtcttcaacc ggcctatcct 480
ctttgacatc gtgtcccggg gctccactgc tgacctggac gggctgctcc cattcttgct 540
gaccacaag aaacgcctaa ctgatgagga gtttcgagag ccatctacgg ggaagacctg 600
cctgcccagg gccttgctga acctgagcaa tggccgcaac gacaccatcc ctgtgctgct 660
ggacatcgcg gagcgaccg gcaacatgag ggagttcatt aactcgccct tccgtgacat 720
ctactatcga ggtcagacag ccctgcacat cgccattgag cgctcgctgca aacatacgt 780
ggaacttctc gtggcccagg gagctgatgt ccacgcccag gcccggtggc gcttcttcca 840
gcccaggat gaggggggct acttctactt tggggagctg cccctgtcgc tggctgcctg 900
caccaaccag ccccatattg tcaactacct gacggagaac cccacaaga aggggacatg 960
cggcgccagg actcgcgagg caacacagtg ctgcatgcgc tgggtggccat tgctgacaac 1020
accgtgaga acaccaagtt tgttaccaag gtgctcaaca acgacggcct ctgcccctc 1140
cgctcttcc ccgacagcaa cctggaggcc gtgctcaaca agcacatcat ccggcgggag 1200
atgatgatgg ctgccaagac gggcaagatt gggatctttc agcacatcat ctatgggcca 1260
gtgacggatg aggacacacg gcacctgtcc cgcaagttca aggactgggc ctatgggcca 1260
gtgtattcct cgctttatga cctctcctcc ctggacacgt gtggggaaga ggcctccgtg 1320
ctggagatcc tgggtgtaca cagcaagatt gagaaccgcc acgagatgct ggctgtggag 1380
cccatcaatg aactgctgcg ggacaagtgg cgcaagttcg gggccgtctc cttctacatc 1440
aacgtggtct cctacctgtg tgccatggtc atcttctact tcaccgocct ctaccgocg 1500
ctggagggca caccgcccga cccttaccgc accacgggtg actacctgcg gctggctggc 1560
gaggtcatta cgctcttcac tggggctcctg ttcttcttca ccaacatcaa agacttggtc 1620
atgaagaaat gccctggagt gaattctctc ttcatatgat gctccttcca gctgctctac 1680
ttcatctact ctgtcctggt gatcgtctca gcagccctct acctggcagg gatcgaggcc 1740
tacctggccg tgatggtctt tgccctggtc ctgggctgga tgaatgccct ttacttcacc 1800
cgtgggctga agctgacggg gacctatagc atcatgatcc agaagattct cttcaaggac 1860
cttttccgat tctgtctcgt ctacttgctc ttcatgatcg gctacgcttc agccctggtc 1920
tccctcctga acccgtgtgc caacatgaag gtgtgcaatg aggaccagac caactgcaca 1980
gtgcccactt acccctcgtg ccgtgacagc gagaccttca gcaccttct cctggacctg 2040
tttaagctga ccatcgcat gggcgacctg gagatgctga gcagcacaa gtaccccgctg 2100
gtcttcatca tctgtctggt gacctacatc atcctcactt ttgtgctgct cctcaacatg 2160
ctcattgccc tcatgggcga gacagtggc caggctcca aggagagcaa gcacatctgg 2220
aagctgcagt gggccaccac catcctggac attgagcgt ccttccccgt attcctgagg 2280
aaggccttcc gctctgggga gatggtcacc gtgggcaaga gctcggacgg cactcctgac 2340
cgcaggtggt gcttcagggg ggatgaggtg aactggctct actggaacca gaacttgggc 2400
atcatcaacg aggaccggg caagaatgag acctaccagt attatggctt ctgcatacc 2460
gtgggcccgc tccgcaggga tcgctggtcc tcggtggtac cccgcgtggt ggaactgaac 2520
aagaactcga acccgacga ggtggtggtg cctctggaca gcacgggaa ccccgctgc 2580
gatggccacc agcagggtta ccccgcaag tggaggactg atgacgccc gctcaggga 2640
ctgcagccca gcccagctt ctctgccac tcatttctag tccagccgca tttcagcag 2700
gccttctggg gtgtccccc acaccctgct tttggcccag aggcgaggga ccagtggagg 2760
ttcaaggagg cccaagaacc tgtggtcccc tggctctgct tcccaacctg g 2811
```

<210> 38  
<211> 2074  
<212> DNA  
<213> Homo sapiens



&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 171968CB1

&lt;400&gt; 38

tggaccccgga	ttctcacctg	gactccaaaa	gctatcttga	cctactggca	tctctgaccc	60
aaatcttaat	tgccccatc	gccctctaca	ccccccagc	actgaccctc	accaggactc	120
cagccccaat	tccatcccaa	atctgtgtag	catctgcttc	tgccgattct	aagagcccta	180
gcacctgcca	agtcccccca	ttaccacact	tcccacactc	agaagcctct	ttggtgggat	240
gctaattggga	aggagtcttg	cctctctgga	ggcaggaggg	gctggccttg	tgccccctcg	300
ggcctctgag	aggtgggagc	aggagaacag	cactcacgag	gggacctcct	tcacctggg	360
aaaggggtgt	ttcttttgta	tttcacagtc	acaggctgaa	tccttcactt	ggccctgccc	420
accgtacagg	tatgctcact	gccggttta	gggaggccag	aaaccaacct	gctcctgcaa	480
aaagaatcca	ggcttggtct	gagtgccctg	tgtaggccag	gcaagttggt	cactgttgca	540
tgagggggcag	tgectctcac	tcttgggctc	gatgccaaag	gaggtggcct	gtcccggctg	600
catgcagaca	tcctggccat	cccagccaca	catgcacgtg	agaggctggg	tgccggcagg	660
gttctctgag	gactggaaga	tgtggccccc	tgectgcctc	cttctctctg	tgaatataag	720
gggcccagttc	ccagcccaaa	gccccaccgc	gggcccctcat	gtttcatcac	caacaggcct	780
actgtctggc	tccttttgac	ctcatcaaag	tcgggctaca	aaaccagaca	gagccaaggg	840
ccagcccagg	gagcccccca	ccccggtaac	agggggccctg	gcactgtgca	gcctccatct	900
tccgggagga	ggggccccgg	gggctgttcc	gaggagcctg	ggccctgacg	ctgagggaca	960
ccccacgggt	ggggatctac	ttcatcacct	atgaagggtc	ctgtcgccag	tacacaccag	1020
aaggccagaa	tcccagctca	gccacgggtg	tggtggcagg	gggctttgca	ggcattgctt	1080
cctgggtggc	agccacgccc	ttagacgtga	tcaagtcccg	gatgcagatg	gatggactga	1140
gacgcagagt	gtaccagggg	atgctggact	gcatggtgag	cagcatccgg	caggaaggac	1200
tgggagtctt	cttcggggg	gtcaccatca	acagtgcctg	cgcttttccc	gtcaatgctg	1260
tcaccttctt	cagctacgaa	tatctctctc	gctggtgggg	atgagccctg	cggcaatgcc	1320
agcagctccc	catcaggccc	acggcctgga	ggccagtttg	agattggagg	ccaggttgaa	1380
agcttgcaaa	tcagtgcagg	aggctcagcc	cttcttaacc	aaggtgcctc	ccaccgcgcg	1440
agatctgggc	tgggcagaca	cctgtgggag	ccggaagcca	gggggcctgt	gcagcctccc	1500
tgtgtagctg	gccttgactc	ctttgcctcc	cacatctgtg	aaacaggag	catgaggcac	1560
aagtgcagctg	gcaagtgggtg	ctggtgacat	cccagctcct	gtcctgtgcc	ttcacctctt	1620
tttttttttt	tttttttttt	ggggagaggg	ggaggtcttt	ctctcttggt	ccccaggggg	1680
cgtgggatct	gcagggggcg	gtgagactcc	tcgggttcat	tggaacacct	cggcggtttc	1740
cacctttccg	gggtctcaag	cgcaattctt	cctggcctta	agccttccca	aggtacgctg	1800
ggagacatat	tagcgcgggc	cggcaacaca	aaccaggtg	taaatttttg	ggtattttta	1860
aggctaagaa	gacaggggtt	taccccattg	tcagcgccag	ggttgggtcc	tggagatctc	1920
tctggatctt	tggggagatc	cggcccggtg	ggtgggcctt	ctcaagagtt	gcgcggggag	1980
attacacagg	gctgtgagag	gccccaccgc	ggcgccccgg	ggtggcctta	cactcttctt	2040
aagggagcct	cggaggactc	ccttttttga	aagg			2074

&lt;210&gt; 39

&lt;211&gt; 1340

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 257274CB1

&lt;400&gt; 39

aggaacacgc	ccatgtgctc	cgggctcctg	gagctcctgc	tgcccatctg	gctctcctgg	60
accctgggga	cccagggtc	tgagccccgc	agtgtgaacg	atcccgggaa	catgtccttt	120
gtgaaggaga	cgggtggaca	gctgttgaaa	ggctacgaca	ttcgccctaag	acccgacttc	180
gggggtcccc	cgggtctgct	ggggatgaac	atcgacatcg	ccagcatcga	catggtttcc	240
gaagtcaaca	tgagattctg	gctgcaggaa	aggggaacga	agacagtggg	ctgtgcgttc	300
caggggtgct	tctgcggtt	ttccaaggct	gctcctgga	ctgggagacc	cgggccccgc	360
accgccagtc	tctgtccgag	gtgctgacag	gcttccgctg	cttcagagag	cgggaggccg	420
cgcctaggcg	ggctctcaga	ggagccgctc	taccgggaga	gtcggaggct	ggtgacccag	480
tgctacttag	gtcttctgtg	aatgcagact	ggattcaata	ttctgacctg	tggaagcgg	540
aggtcagtac	cccagaggtg	gaagcgggct	tttgccagga	gtgcttttag	acgccaggga	600
atcaggagaa	ggatggccct	ttcatattgt	aattgagctg	aaacgcgctg	ggatttgaaa	660
actagtttag	ttttctgcgg	agggacactc	tggaaggagc	atgtgtaaac	aatatttgtt	720

```

tttttggaga aattgtttgg caacttttct ttccggacata caacattgag aatacagtg 780
gacatgggtt tagatccact ctgtaggctt ctttaactct gtgtacgtcg gaatcacgtg 840
ggcaaaacttg ctttaaatgc gtgcctgcta gttcctcact ccaagggatt ctgatcaaga 900
taggctggag tcttgggggc cctgcattta aaactacttc ttcaaataatt ttgaagcggg 960
tattttgtgg acaatatttt gagaaacttt gcagtaagtc acagatatat gaatatatac 1020
aaataaagta aaactatttt taaaaataga cgttagttag atacttttaa aatacctacc 1080
actaaggatc ttataggaag aaacttttga ccacgccaaag ctgtttctcat taattttttca 1140
ctgttaactc aaagctttta aatttacaat cccatgtatt taaaaatgta cttttttcca 1200
aagtatatca cttaggacat ttgttaagtc aatattgtat cagttaaagt gttagcaagg 1260
aacacagaag gaatgtgact gcataaatta ccttacagta aaaattaaca tgtctatttt 1320
actttttatg taatacat 1340

```

&lt;210&gt; 40

&lt;211&gt; 6027

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6355991CB1

&lt;400&gt; 40

```

atggagcaaa cagtgcctgt accaccagga cctgacagct tcaacttctt caccagagaa 60
tctcttgcgg ctattgaaag acgcattgca gaagaaaagg caaagaatcc caaaccagac 120
aaaaaagatg acgacgaaaa tggcccaaag ccaaatagtg acttggaagc tggaaagaac 180
cttccattta tttatggaga cattcctcca gagatgggtg cagagccctt ggaggacctg 240
gaccctactc atatcaataa gcagactttt atagtattga ataaagggaa ggccatcttc 300
cggttcagtg ccacctctgc cctgtacatt ttaactccct tcaatcctct taggaaaata 360
gctattaaga ttttgggtaca ttcattatct agcatgctaa ttatgtgcac tattttgaca 420
aactgtgtgt ttatgacaat gagtaaccct cctgattgga caaagaatgt agagtacacc 480
ttcacaggaa tatatacttt tgaatcactt ataaaaatta ttgcaagggg attctgttta 540
gaagatttta ctttcccttcg ggatccatgg aactggctcg atttcaactgt cattacattt 600
gcgtacgtca cagagtttgt ggacctgggc aatgtctcgg cattgagaac attcagagtt 660
ctccgagcat tgaagacgat ttcagtcatt ccaggcctga aaaccattgt cctgagcgtg 720
atccagtcctg tgaagaagct ctcagatgta atgacctga ctgtgttctg tctgagcgtg 780
tttgctctaa ttgggctgca gctgttcatg ggcaacctga ggaataaatg tatacaatgg 840
cctcccacca atgcttccct ggaggaacat agtatagaaa agaataaac tgtgaattat 900
aatggtacac ttataaatga aactgtcttt gagtttgact ggaagtcata tattcaagat 960
tcaggatata attatttccct ggagggtttt ttagatgcac tactatgtgg aaatagctct 1020
gatgcagggc aatgtccaga gggatatatg tgtgtgaaag ctggtagaaa tccaattat 1080
ggctacacaa gctttgatac cttcagtttg gcttttttgt ccttgtttcg actaatgact 1140
caggacttct gggaaaatct ttatcaactg acattacgtg ctgctgggaa aacgtacatg 1200
atattttttg tattgggtcat tttcttgggc tcattctacc taataaattt gatcctggct 1260
gtggtggcca tggcctacga ggaacagaat caggccacct tggagaagc agaacagaaa 1320
gaggccgaat ttcagcagat gattgaacag cttaaaaagc aacaggaggc agctcagcag 1380
gcagcaacgg caactgcctc agaacattcc agagagccca gtgcagcagg caggctctca 1440
gacagctcat ctgaagcctc taagttagt tccaagagtg ctaaggaaaag aagaaatcgg 1500
aggaagaaaa gaaaacagaa agagcagctc ggtggggaag agaaagatga ggatgaattc 1560
caaaaatctg aatctgagga cagcatcagg aggaaaggtt ttcgcttctc cattgaaggg 1620
aaccgattga catatgaaaa gaggtactcc tccccacacc agtctttgtt gagcatccgt 1680
ggctccctat tttcaccaag gcgaaatagc agaacaagcc ttttcagctt tagagggcga 1740
gcaaaggatg tgggatctga gaacgacttc gcagatgatg agcacagcac ctttgaggat 1800
aacgagagcc gtagagattc cttgttttgt ccccgacgac acggagagag acgcaacagc 1860
aacctgagtc agaccagtag gtcacccgg atgctggcag tgtttccagc gaatggggaag 1920
atgcacagca ctgtggattg caatggtgtg gtttcccttg ttggtggacc ttcagttcct 1980
acatcgcttg ttggacagct tctgccagag gtgataatag ataagccagc tactgatgac 2040
aatggaacaa ccactgaaac tgaaatgaga aagagaaggt caagttcttt ccacgtttcc 2100
atggactttc tagaagatcc ttcccaaagg caacgagcaa tgagtatagc cagcattcta 2160
acaaatacag tagaagaact tgaagaatcc aggcagaaat gccaccctg ttggtataaa 2220
ttttccaaca tattcttaat ctgggactgt tctccatatt ggttaaaagt gaaacatggt 2280
gtcaacctgg ttgtgatgga cccatttgtt gacctggcca tcaccatctg tattgtctta 2340
aatactcttt tcatggccat ggagcactat ccaatgacgg accatttcaa taatgtgctt 2400
acagtaggaa acttgggtatt cactgggact tttacagcag aaatgtttct gaaaattatt 2460

```

gccatggatc	cttactatta	tttccaagaa	ggctggaata	tctttgacgg	ttttattgtg	2520
acgcttagcc	tggtagaact	tggtactgcc	aatgtggaag	gattatctgt	tctccgttca	2580
tttcgattgc	tgcgagtttt	caagttggca	aaatccttggc	caacgttaaa	tatgctaata	2640
aagatcatcg	gcaattccgg	gggggctctg	ggaaatttaa	ccctcgtctt	ggccatcatc	2700
gtcttcattt	ttgcccgtgt	cggcatgcag	ctctttggta	aaagctacaa	agattgtgtc	2760
tgcaagatcg	ccagtgattg	tcaactccca	cgtgtggcaca	tgaatgactt	cttccactcc	2820
ttcctgattg	tggtccgcgt	gctgtgtggg	gagtggatag	agaccatgtg	ggactgtatg	2880
gaggttgctg	gtcaagccat	gtgccttact	gtcttcatga	tggtcatggg	gattggaaac	2940
ctagtggtag	tgaatctctt	tctggccttg	cttctgagct	catttagtgc	agaacaacct	3000
gcagccactg	atgatgataa	tgaatgaat	aatctccaaa	ttgctgtgga	taggatgcac	3060
aaaggagtag	cttatgtgaa	aagaaaaata	tatgaattta	ttcaacagtc	cttcattagg	3120
aaacaaaaga	ttttagatga	aattaaacca	cttgatgata	taaacaacaa	gaaagacagt	3180
tgtatgtcca	atcatacagc	agaaattggg	aaagatcttg	actatcttaa	agatgtaaat	3240
ggaactacaa	gtggtatagg	aactggcagc	agtgttgaaa	aatacattat	tgatgaaagt	3300
gattacatgt	cattcataaa	caaccccagt	cttactgtga	ctgtaccaat	tgctgtagga	3360
gaatctgact	ttgaaaattt	aaacacggaa	gacttttagta	gtgaatcggg	tctggaagaa	3420
agcaagagag	aactgaatga	aagcagtagc	ctatcagaag	gtagcactgt	ggacatcggc	3480
gcacctgtag	aagaacagcc	cgtagtggaa	cctgaagaaa	ctcttgaacc	agaagcttgt	3540
ttcactgaag	gttgtgtaca	aagattcaag	tggtgtcaaa	tcaatgtgga	agaaggcaga	3600
ggaaaacaat	ggtggaacct	gagaaggacg	tggttccgaa	tagttgaaca	taactgggtt	3660
gagaccttca	ttgttttcat	gattctcctt	agtagtggtg	ctctggcatt	tgaagatata	3720
tatattgatc	agcgaaaagac	gattaagacg	atggttgaat	atgctgacaa	ggttttctact	3780
tacattttca	ttctggaaat	gcttctaaaa	tgggtggcat	atggctatca	aacatatttc	3840
accaatgcct	ggtgttggct	ggacttctta	attgttgatg	tttcatgtgt	cagtttaaca	3900
gcaaatagct	tggtgtactc	agaacttgga	gccatcaaat	ctctcaggac	actaagagct	3960
ctgagacctc	taagagcctt	atctcgattt	gaagggatga	gggtagtgtg	gaatgccctt	4020
ttaggagcaa	ttccatccat	catgaatgtg	cttctgggtt	gtcttatatt	ctggctaatt	4080
ttcagcatca	tgggcgtaaa	ttgttttgc	ggcaaatctt	accactgtat	taacaccaca	4140
actggtgaca	ggtttgacat	cgaagacgtg	aataatcata	ctgattgcct	aaaactaata	4200
gaaagaaatg	agactgctcg	atggaaaaat	gtgaagttaa	actttgataa	tgtaggattt	4260
gggtatctct	ctttgcttca	agttgccaca	ttcaaaggat	ggatggatat	aatgtatgca	4320
gcagttgatt	ccagaaatgt	agaactccag	cctaagtatg	aagaaagtct	gtacatgtat	4380
ctttactttg	ttattttcat	catctttggg	tccttcttca	ccttgaacct	gtttattggg	4440
gtcatcatag	ataatttcaa	ccagcagaaa	aagaagtttg	gaggtcaaga	catctttatg	4500
acagaagaac	agaagaaata	ctataatgca	atgaaaaaat	taggatcgaa	aaaaccgcaa	4560
aagcctatac	ctcgaccagg	aaacaaattt	caaggaaatg	tctttgactt	cgtaaccaga	4620
caagtttttg	acataagcat	catgattctc	atctgtctta	acatggtcac	aatgatgggtg	4680
gaaacagatg	accagagtga	atatgtgact	accattttgt	cacgcatcaa	tctgggtgtc	4740
attgtgctat	ttactggaga	gtgtgtactg	aaactcatct	ctctacgcca	ttattatttt	4800
accattggat	ggaatatatt	tgattttgtg	gttgtcatte	tctccattgt	aggtatgttt	4860
cttgccgagc	tgatagaaaa	gtatttctgt	tccctacccc	tggtccgagt	gatccgtctt	4920
gctaggattg	gcogaatcct	acgtctgact	aaaggagcaa	aggggatccg	cacgctgtct	4980
tttgctttga	tgatgtccct	tctgcgttg	tttaacatcg	gcctcctact	cttcctagtc	5040
atgttcatct	acgccatctt	tgggatgtcc	aactttgcct	atgttaagag	ggaagttggg	5100
atcgatgaca	tgttcaactt	tgagaccttt	ggcaacagca	tgatctgcct	attccaaatt	5160
acaacctctg	ctggctggga	tggattgcta	gcacccattc	tcaacagtaa	gccacccgac	5220
tgtgacctta	ataaagttaa	ccctggaagc	tcagttaagg	gagactgtgg	gaacccatct	5280
gttggaaattt	tcttttttgt	cagttacatc	atcatatctg	tctgggttgt	ggtgaacatg	5340
tacatcgctg	tcactctgga	gaacttcagt	gttgctactg	aagaaaagtgc	agagcctctg	5400
agtgaggatg	acttttgagat	gttctatgag	gtttgggaga	agtttgatcc	cgatgcaact	5460
cagttcatgg	aatttgaaaa	attatctcag	tttgcagctg	cgcttgaacc	gcctctcaat	5520
ctgccacaac	caaacaaact	ccagctcatt	gccatggatt	tgcccatggg	gagtggtgac	5580
cggatccact	gtcttgatat	cttatttgc	tttacaagag	gggttctagg	agagagtggg	5640
gagatggatg	ctctacgaat	acagatggaa	gagcgattca	tggcttccaa	tccttccaag	5700
gtctctatc	agccaatcac	tactacttta	aaacgaaaac	aagagggaagt	atctgtgtgc	5760
attattcagc	gtgcttacag	acgccacctt	ttaaagcgaa	ctgtaaaaca	agcttctctt	5820
acgtacaata	aaaacaaaat	caaaggtggg	gctaattctt	ttataaaaga	agacatgata	5880
attgacagaa	taaatgaaaa	ctctattaca	gaaaaaactg	atctgaccat	gtccactgca	5940
gcttgtccac	cttcctatga	ccgggtgaca	aagccaattg	tggaaaaaca	tgagcaagaa	6000
ggcaagatg	aaaaagccaa	agggaaa				6027

&lt;210&gt; 41

&lt;211&gt; 2168

<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 70035348CB1

<400> 41  
attagctttg cccgaagttt ttccccacac tcttctttag catgctatta tggggaaagt 60  
gaccactcct gggagcgggg gtgggtcggg cggtttggtg gcggggaagc ggctgtaact 120  
tctacgtgac catggtacct gttgaaaaca ccgagggccc cagtctgctg aaccagaagg 180  
ggacagccgt ggagacggag ggcagcggca gccggcatcc tccctgggcg agaggctgcy 240  
gcatgtttac cttcctgtca tctgtcactg ctgctgtcag tggcctcctg gtgggttatg 300  
aacttgggat catctctggg gctcttcttc agatcaaaac cttattagcc ctgagctgcc 360  
atgagcagga aatggttggt agctccctcg tcattggagc cctccttgcc tcaactaccg 420  
gaggggtcct gatagacaga tatggaagaa ggacagcaat catcttgta tctgctgc 480  
ttggactcgg aagcttagtc ttgatcctca gtttatccta caggttctt atagtgggac 540  
gcattgccat aggggtctcc atctccctct cttccattgc cacttggtgt tacatcgag 600  
agattgctcc tcaacacaga agaggccttc ttgtgtcact gaatgagctg atgattgtca 660  
tcggcattct ttctgcctat atttcaaatt acgcatttgc caatgttttc catggctgga 720  
agtacatggt tggcttggtg attcccttgg gagttttgca agcaattgca atgtattttc 780  
ttcctccaag cctcgggttt ctggtgatga aaggacaaga gggagctgct agcaagggtc 840  
ttggaagggt aagagcactc tcagatacaa cactgtgate aaatcctccc 900  
tgaaagatga atatcagtac agtttttggg atctgtttcg ttcaaaagac aacatgcgga 960  
cccgaataat gataggacta acactagtat tttttgtaca aatcactggc caaccaaaaca 1020  
tattgttcta tgcatacaat gttttgaagt cagttggatt tcaaagcaat gaggcagcta 1080  
gcctcgctc cactgggggt ggagtcgtca aggtcattag caccatccct gccactcttc 1140  
ttgtagacca tgtcggcagc aaaacattcc tctgcattgg ctctctgtg atggcagctt 1200  
cgttggtgac catgggcatc gtaaatctca acatccacat gaacttcacc catatctgca 1260  
gaagccacaa ttctatcaac cagtccttgg atgagtctgt gatttatgga ccaggaaacc 1320  
tgtcaaccaa caacaatact ctcagagacc acttcaaagg gatctcttc catagcagaa 1380  
gctcactcat gcccctgaga aatgatgtgg ataagagagg ggagacgacc tcagcatcct 1440  
tgctaaatgc tggattaagc cacactgaat accagatagt cacagaccct ggggacgtcc 1500  
cagctttttt gaaatggctg tccttagcca gcttgctgtg ttatgttgct gctttttcaa 1560  
ttggtctagg accaagagat gttatcttta tcggacagtc aacaaacttg cctctgctc 1620  
cagagggtga cactatctct atctccaaga ctatttatta tgcagcctac aacaaggcta 1680  
ttatacaaac agccttgga agacagccta gagcaaagac agtcagtgcc ttttcccata 1740  
agacatgaag aaatgtgaga gacctacgga gaactggctc ccaacccaaa tctcctaaaa 1800  
ctcaaagtgc tttctttcta ttcgaaacaa caaactagaa ttttgaaaaa ctcaaagacc 1860  
atagagccta gctttttgct ctgtttggtt ttatggagct gaaccagcct attggagggt 1920  
gggtatcaat gttggaagca tggatcatct gccgtgaaat ttaaacttag atttaacaa 1980  
ataactctgg ctcttaaaaa ttttgttcat tggatatattg cacagctaaa gattatgaca 2040  
gctccaagga tgtggagcag caggttctaa tttggaagtt tacctagtgg cttcatttca 2100  
agacctactg ggtttaaggc aaagaggctg acattgcaga agcacagggtg tttcaaatca 2160  
gattctgg 2168

<210> 42  
<211> 2229  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7472539CB1

<400> 42  
atggaatacc aggcgtccga ggtgatcggg cagcgtcagt cttcagccac taagccagga 60  
agatctggga aggagtcagt cacagagccc tgggcccagag ttccaggggc tctgggagtg 120  
gctgccaggc agatgcacc caagtcaata atcacattca gagagataaa tggggagtag 180  
actggggctg tggattttcc caggctagga gtcctgctt ctgaggaaac agcgtcaga 240  
gagctgaaga tgagcaagga gctggcagca atggggcctg gagcttcagg ggaacggggtc 300  
aggactgaga cagctccaca catagcactg gactccagag ttggtctgca cgcctacgac 360  
atcagcgtgg tggatcatca ctttgtcttc gtcattgctg tggggatctg gtcgtccatc 420

cgtgcaagtc	gagggacat	tggcggctat	ttcctggccg	ggagttggag	catctctgat	480
gtccagcaat	gtgggcagtg	gcttggtcat	cgccctggct	gggacagggg	ctgccggagg	540
ccttgccgta	ggtggcttcg	agtggaaactg	ctcctggccc	ttggctgggt	cttcgtccct	600
gtgtacatcg	cagcaggtgt	ggtcacaatg	ccgcagtatc	tgaagaagcg	atttgggggc	660
cagaggatcc	aggtgtacat	gtctgtcctg	tctctcatcc	tctacatctt	caccaagatc	720
tcgactgaca	tcttctctgg	agccctcttc	atccagatgg	cattgggctg	gaacctgtac	780
ctctccacag	ggatcctgct	ggtggtgact	gccgtctaca	ccattgcagg	tggcctcatg	840
gccgtgatct	acacagatgc	tctgcagacg	gtgatcatgg	tagggggagc	cctggtcctc	900
atgtttctgg	gctttcagga	cgtgggctgg	taccagggcc	tggagcagcg	gtacaggcag	960
gccatcccta	atgtcacagt	ccccaacacc	acctgtcacc	tcccacggcc	cgatgctttc	1020
cacattcttc	gggacctgt	gagcggggac	atcccttggc	caggtctcat	tttcgggctc	1080
acagtgtg	ccacctgggtg	ttggtgcaca	gaccaggtca	ttgtgcagcg	gtctctctcg	1140
gccaaagtc	tgtctcatgc	caaggaggc	tccgtgctgg	ggggctacct	gaagatcctc	1200
cccatgttct	tcctcgtcat	gcctggcatg	atcagccggg	ccctgttccc	agacgaggtg	1260
ggctgcgtgg	acctgatgt	ctgccaaaga	atctgtgggg	cccagtgagg	atgttccaac	1320
attgcctacc	ctaagttggt	catggccctc	atgcctgttg	gtctgcgggg	gctgatgatt	1380
gccgtgatca	tggccgctct	catgagctca	ctcacctcca	tcttcaacag	cagcagcacc	1440
ctgttcacca	ttgatgtgtg	gcagcgcttc	cgcaggaagt	caacagagca	ggagctgatg	1500
gtggtgggca	gagtgtttgt	ggtgttccctg	gttgatcatca	gcacccctctg	gatccccatc	1560
atccaaagct	ccaacagtgg	gcagctcttc	gactacatcc	aggctgtcac	cagttacctg	1620
gccccaccca	tcaccgctct	cttctgtctg	gccatcttct	gcaagagggt	cacagagccc	1680
ggagctttct	ggggccctcgt	gtttggcctg	ggagtggggc	ttctgcgtat	gatcctggag	1740
ttctcatacc	cagcgccagc	ctgtggggag	gtggaccgga	ggccagcagt	gctgaaggac	1800
ttccactacc	tgtactttgc	aatcctcttc	tgcgggctca	ctgccatcgt	cattgtcatt	1860
ctcacacgcc	tcacatgggtg	gactcggaa	tgccccctct	ctgagctgga	gaaggaggcc	1920
cacgagagca	caccggagat	atccgagagg	ccagccgggg	agtgccttgc	aggaggtgga	1980
gcggcagaga	actcgagcct	gggccaggag	cagcctgaag	ccccaagcag	gtcctgggga	2040
aagttgctct	ggagctgggt	ctgtgggctc	tctggaacac	cggagcaggc	cctgagccca	2100
gcagagaagg	ctgcgctaga	acagaagctg	acaagcattg	aggaggagcc	actctggaga	2160
catgtctgca	acatcaatgc	tgtccttttg	ctggccatca	acatcttctc	ctggggctat	2220
tttgcgtga						2229

&lt;210&gt; 43

&lt;211&gt; 1520

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 817477CB1

&lt;400&gt; 43

gcgcctcgtc	gggoccttcc	tctctacctg	cctctccaac	ccctctcggc	cccagaccac	60
ccggcagcgg	gggtgggtgt	gcagaggtgc	ggcgtccaga	accccggtc	ctgcagaggc	120
tctgggtggc	agcagccctg	ttaccgctta	gatggcgcg	aggacagagc	ccccgcagcg	180
gggctgggga	tgggtgggtg	tgtctcagc	gttcttcag	tcggcgcttg	tgtttggggg	240
gctccgctcc	tttggggtct	tcttcgtgga	gtttgtggcg	gcgtttgagg	agcaggcagc	300
gcgcgtctcc	tggatcgcc	ccataggaat	cgcggtgcag	cagtttggga	gcccggtagg	360
cagtgccttg	agcacgaagt	tccggcccag	gcccgtgggtg	atgactggag	gcactcttggc	420
tgcgctgggg	atgctgctcg	cctcttttgc	tacttccctg	acccacctat	acctgagtat	480
tgggttgctg	tcaggctctg	gctgggcttt	gaacctcgct	ccgaccttgg	cctgcctgtc	540
ctgttatattc	tctcgccgac	gatccctggc	caccgggctg	gcactgacag	gcgtgggccc	600
ctcctccttc	acatttgccc	cctttttcca	gtggctgctc	agccactacg	cctggagggg	660
gtccctgctg	ctgggtgtctg	ccctctccct	ccacctagtg	gectgtgggtg	ctctctctcg	720
cccacccctcc	ctggctgagg	acctgtgtgt	gggtggtccc	agggcccaac	tcacctctct	780
cctccatcat	ggcccttccc	tcggttacac	gtgtgcccctc	acctgatca	acactggcta	840
cttcattccc	tacctccacc	tgggtggccc	tctccaggac	ctggattggg	acccactacc	900
tgtctgccttc	ctactctcag	ttgttgctat	ttctgacctc	gtggggcggtg	tgggtctccgg	960
atggctggga	gatgcagtc	cagggcctgt	gacacgactc	ctgatgctct	ggaaccacctt	1020
gactgggggtg	tcactagccc	tgttccctgt	agctcaggct	cccacagccc	tgggtggctct	1080
ggctgtggcc	tacggcttca	catcaggggc	tctggcccga	ctggccttct	ctgtgtgccc	1140
tgaactaata	gggactagaa	ggatttactg	tggcctggga	ctgttgacaga	tgatagagag	1200
catcgggggg	ctgctggggc	ctcctctctc	aggctacctc	cgggatgtga	caggcaacta	1260

cacggcttct tttgtggtgg ctggggcctt ccttctttca gggagtggca ttctcctcac 1320  
cctgccccac ttcttctgct tctcaactac tacctccggg cccaggacc ttgtaacaga 1380  
agcactagat actaaagtgc cctaccctaa ggagggactg gaaggaggact gaactccaca 1440  
gagtcaggcc cagaaagcca aagcttgaca gctccaggtc ttctcttgcc acgtcttggg 1500  
ctccacagaa ccacagtgc 1520

&lt;210&gt; 44

&lt;211&gt; 3950

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1442166CB1

&lt;400&gt; 44

gccagcctgt tctgttgccc tggtcttctc tagtccagge tgccatggcg gcgctcaggg 60  
cttaccggaa gtaaaacttc ggaagtggag cgttctcttg cccggaagtg agcgcgggcg 120  
taggaaagat ggcggcagcg gcggcggttg gcaacgcggg gccctgcggg gcccgccctt 180  
gcgggggtccg gcttgacggg cagccccagc ccggggccgca gccgcgcgcg ctccttgccg 240  
ccggggccggc gctcatagcg aacgggtgac agctgggtggc tgccgtgtgg ccgtaccggc 300  
gggtggcgct gttgcggcgc ctcacgggtg tgccattcgc cgggctgctt taccggcgct 360  
gggtgggtgc cgcagccgct ggctgctggg gctggggcag cagttgggtg cagatccccg 420  
aagctgcgct gctcgtgctt gccaccatct gcctcgcgca cgcgctcact gtcctctcgg 480  
ggcattgggtc tgtgcacgcg cattgcgcgc tcacctgcac cccggagtac gaccccagca 540  
aagcgacctt tgtgaagggt gtgccaaccc ccaacaatgg ctccacggag ctctggtggc 600  
tgaccgcgaa tgagggcgaa gacgggcttg aggtgctgtc cttcgaattc cagaagatca 660  
agtattccta cgatgccctg gagaagaagc agtttctccc cgtggccttt cctgtgggaa 720  
acgccttctc atactatcag agcaacagag gcttccagga agactcagag atccgagcag 780  
ctgagaagaa atttgggagc aacaaggccg agatgggtgg gcctgacttc tcggagcttt 840  
tcaaggagag agccacagcc ccttctcttg tatttcagggt gttctgtgtg gggctctggg 900  
gcctggatga gtactggtac tacagcgtct ttacgctatc catgctggtg gcgttcgagg 960  
cctcgctggt gcagcagcag atgcggaaca tgcctggagat ccggaagatg ggcaacaagc 1020  
cccacatgat ccaggtctac cgaagccgca agtggaggcc cattgccagt gatgagatcg 1080  
taccagggga catcgtctcc atcgccgct cccacagga gaacctggtg ccatgtgacg 1140  
tgcttctgct gcgaggccgc tgcacgtag acgaggccat gctcacgggg gagtcgtagc 1200  
cacagatgaa ggagcccatc gaagacctca gccagaccg ggtgctggac ctccaggctg 1260  
attcccggt gcacgtcatc ttcgggggca ccaaggtggt gcagcacatc cccccacaga 1320  
aagccaccac gggcctgaag ccggttgaca gcgggtgcgt ggccctacgtc ctgcggaccg 1380  
gattcaacac atcccagggc aagctgtgc gcacctctt cttcgggggtc aagaggggtga 1440  
ctgcgaacaa cctggagacc ttcattctca tctcttctc cctggtgttt gccatcgctg 1500  
cagctgccta tgtatggatt gaaggtagca aggacccag ccggaaccgc tacaagctgt 1560  
ttctggagtg caccctgac ctcacctcg tcgtgcctcc tgagctgccc atcgagctgt 1620  
ccctggccgt caacacctcc ctcacgccc tggccaagct ctacatgtac tgcacagagc 1680  
ccttccggt cccctttgct ggcaaggctc aggtgtgctg ctttgacaag acggggacgt 1740  
tgaccagtga cagctggtg gtgcgcgggt tggccgggct gagagacggg aaggaggtga 1800  
ccccagtgtc cagcatccct gttagaaacac accgggccct ggccctcgtc cactcgtc 1860  
tgcagctgga cgacggcacc ctcgtgggtg accctctaga gaaggccatg ctgacggccg 1920  
tggactggac gctgaccaa gatgagaaag tattcccccg aagtattaaa actcaggggc 1980  
tgaaaattca ccagcgcttt cattttgcca gtgccctgaa gcgaatgtcc gtgcttgcc 2040  
cgtatgagaa gctgggctcc accgacctct gctacatcgc ggccgtgaag ggggcccccg 2100  
aaactctgga ctccatgttc tcccagtgc cggccgacta ccaccacatc cacaccgaga 2160  
tctcccggga aggagccgc gtcctggcgc tggggtagaa ggagctggga cactcactc 2220  
accagcaggc ccgggaggtc aagcgggagg cctggagtg cagcctcaag ttcgtcggct 2280  
tcattgtggt ctctgcccg ctcaaggctg actccaaggc cgtgatccgg gagatccaga 2340  
atgcgtccca ccgggtggtc atgatcacgg gagacaaccc gctcactgca tgccacgtgg 2400  
cccaggagct gcacttcatt gaaaaggccc acacgctgat cctgcagcct cctccgaga 2460  
aaggccggca gtgcgagtgg cgtccattg acggcagcat cgtgctgccc ctggcccggg 2520  
gtccccaaa ggcactggcc ctggagtac cactgtgcct cacaggcgac ggcttggccc 2580  
acctgcaggc caccgacccc cagcagctgc tccgctcat ccccatgtg caggtgttcg 2640  
cccgtgtggc tcccaagcag aaggagtttg tcatcaccag cctgaaggag ctgggctacg 2700  
tgaccctcat gtgtggggat ggcaccaacg acgtgggcgc cctgaagcat gctgacgtgg 2760  
gtgtggcgct cttggccaat gccctgagc ggggtgtcga gcggcgacgg cggccccggg 2820

acagcccaac	cctgagcaac	agtggcatca	gagccacctc	caggacagcc	aagcagcgg	2880
cggggctccc	tcctctccgag	gagcagccaa	cctcccagag	ggaccgcctg	agccaggtgc	2940
tgcgagacct	cgaggacgag	agtacgcccc	ttgtgaaact	gggggatgcc	agcatcgag	3000
cacccttcac	ctccaagctc	tcattccatcc	agtgcattctg	ccacgtgatc	aagcagggcc	3060
gctgcacgct	ggtgaccacg	ctacagatgt	tcaagatcct	ggcgtcaat	gccctcatcc	3120
tggcctacag	ccagagcgctc	ctctacctgg	agggagtcaa	gttcagtgc	ttccaggcca	3180
ccctacaggg	gctgctgctg	gccggctgct	tcctcttcat	ctcccgttcc	aagcccctca	3240
agaccctctc	ccgagaacgg	cccctgcccc	acatcttcaa	cctgtacacc	atcctcaccg	3300
tcatgctcca	gttctttgtg	cacttctctga	gccttgtcta	cctgtaccgt	gaggccagg	3360
cccggagccc	cgagaagcag	gagcagttcg	tggacttgta	caaggagttt	gagccaagcc	3420
tggtcaacag	caccgtctac	atcatggcca	tggccatgca	gatggccacc	ttcgccatca	3480
attacaaagg	cccgccttcc	atggagagcc	tggccgagaa	caagcccctg	gtgtggagtc	3540
tggcagtttc	actcctggcc	atcattggcc	tgctcctcgg	ctcctcgccc	gacttcaaca	3600
gccagtttgg	cctcgtggac	atccctgtgg	aggtcctgct	cctggacttc	tgccctggcg	3660
tcctggccga	ccgcgtcctg	cagttcttcc	tggggacccc	gaagctgaaa	gtgccttctc	3720
gagatggcag	tgctggtacc	cactgcccac	cctggctgcc	gctgggcggg	aaccccaaca	3780
gggccccggg	agggaaacct	gcccccaacc	cccacagca	aggctgtaca	gtctcgccct	3840
tggaaagactg	agctgggacc	cccacagcca	tcgctggct	tggccagcag	aaccagcccc	3900
aagccagcac	ctttggtaaa	taaagcagca	tctgagattt	taaaaaaaaa		3950

&lt;210&gt; 45

&lt;211&gt; 5540

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2311751CB1

&lt;400&gt; 45

tctacttctc	ctacggcttc	gtctggatcc	aggacatgat	ggagcgcgcc	atcatcgaca	60
ctttttgtggg	gcacgacgtg	gtggagccag	gcagctacgt	gcagatgttc	ccctaccctc	120
gctacacacg	cgatgacttc	ctggtttgtca	ttgagcacat	gatgccgctg	tgcatgggtga	180
tctcctgggt	ctactccgtg	gccatgacca	tccagcacat	cgtggcggag	aaggagcacc	240
ggctcaagga	ggtgatgaag	accatgggcc	tgaacaacgc	ggtgcactgg	gtggcctgg	300
tcataccggg	ctttgtgcag	ctgtccatct	ccgtgacagc	actcaccgcc	atcctgaagt	360
acggccagggt	gcttatgcac	agccacgtgg	tcattcatctg	gctcttctctg	gcagtctacg	420
cggtggccac	catcatgttc	tgcttctctg	tgctgtgtgt	gtactccaag	gccaagctgg	480
cctcggcctg	cggtggcatc	atctacttcc	tgagctacgt	gccctacatg	tacgtggcga	540
tcgagagga	ggtggcgcat	gataagatca	cggccttcga	gaagtgcate	gcgtccctca	600
tgtccacgac	ggcctttggt	ctgggctcta	agtaactcgc	gctgtatgag	gtggccggcg	660
tgggcatcca	gtggcacacc	ttcagccagt	ccccgttga	gggggacgac	ttcaacttgc	720
tcctggctgt	caccatgctg	atggtggacg	ccgtggtcta	tggcatcctc	acgtggtaca	780
ttgaggctgt	gcacccaggc	atgtacgggc	tggcccgccc	ctgggtacttc	ccactgcaga	840
agtcctactg	gctgggcagt	gggcggacag	aagcctggga	gtggagctgg	ccgtgggcac	900
gcaccccccg	cctcagtgtc	atggaggagg	accaggcctg	tggcatggag	agcggcgct	960
ttgaggagac	ccgtggcatg	gaggaggagc	ccaccacct	gcctctggtt	gtctgcgtgg	1020
acaaactcac	caaggtctac	aaggacgaca	agaagctggc	cctgaacaag	ctgagcctga	1080
acctctacga	gaaccagggt	gtctccttct	tgggcccaca	cggggcgggc	aagaccacca	1140
ccatgtccat	cctgaccggc	ctgttccctc	caacgtcggg	ttccgccacc	atctacgggc	1200
acgacatccg	cacggagatg	gatgagatcc	gcaagaacct	gggcatgtgc	ccgcagcaca	1260
atgtgctctt	tgaccggctc	acggtggagg	aacacctctg	gttctactca	cggctcaaga	1320
gcatggctca	ggaggagatc	cgcagagaga	tggacaagat	gatcgaggac	ctggagctct	1380
ccaacaaacg	gcactcactg	gtgcagacat	tgctcgggtg	catgaagcgc	aagctgtccg	1440
tggccatcgc	gctcgtgggc	ggctctcgcg	ccatcatcct	ggacgagccc	acggcggggc	1500
tggaccccta	cgcgcggcgc	gccatctggg	acctcatcct	gaagtacaag	ccaggccgca	1560
ccatccttct	gtccacccac	cacatggatg	aggctgacct	gcttggggac	cgcattgcca	1620
tcattctcca	tgggaagctc	aagtgtcgcg	gctccccgct	cttctcaag	ggcacctatg	1680
gcgacgggta	ccgcctcacg	ctgggtcaagc	ggcccccgca	gccggggggc	ccccaaagac	1740
cagggctggc	atccagcccc	ccaggtcggg	ccccgctgag	cagctgctcc	gagctccagg	1800
tgtcccagtt	catccgcaag	catgtggcct	ctgcctgct	ggtctcagac	acaagcacgg	1860
agctctccta	catcctgccc	agcgaggccg	ccaagaagg	ggctttcgag	cgcctcttcc	1920
agcacctgga	gcgcagcctg	gatgcactgc	acctcagcag	cttcgggctg	atggacacga	1980



ccctggagga agtgttccctc aaggtgtcgg aggaggatca gtcgctggag aacagtgagg 2040  
ccgatgtgaa ggagtccagg aaggatgtgc tccctggggc ggaggggccc ggcgtctgggg 2100  
agggtcacgc tggcaatctg gcccggtgct cggagctgac ccagtcgcag gcctcgctgc 2160  
agtgcgctc atctgtgggc tctgcccgtg gcgacgaggg agctggctac accgacgtct 2220  
atggcgacta ccgccccctc tttgataacc cacaggaccc agacaatgtc agcctgcaag 2280  
aggtggaggg agaggccctg tcgagggtcg gccagggcag ccgcaagctg gacggcgggg 2340  
ggctgaaggt gcgccagttc cacgggctgc tgggtcaaacg cttccactgc gcccgccgca 2400  
actccaaggg actcttctcc cagatcttgc tgccagcctt cttcgtctgc gtggccatga 2460  
ccgtggccct gtccgtcccg gagattggtg atctgcccc cctggtcctg tcacctccc 2520  
agtaccacaa ctacaccag ccccggtggca atttcatccc ctacgccaac gaggagcgcc 2580  
gcgagtaccg gctgcggcta tcgcccagcg ccagccccc gcagctcgtg agcacgttcc 2640  
ggctgccgtc gggggtgggt gccacctgcg tgetcaagtc tcccggcaac ggctcgctgg 2700  
ggccccagtt gaacctgagc agcggggagt cgcgcctgct ggcggtcgg ttcttcgaca 2760  
gcatgtgtct ggagtccttc acacaggggc tgccactgtc caatttcgtg ccacccccac 2820  
cctcgcccgc cccatctgac tcgccagcgt ccccgagatga ggacctgcag gcctggaacg 2880  
tctccctgcc gccaccgct gggccagaaa tgtggacgtc ggcaccctcc ctgcgcgccc 2940  
tggtacggga gcccgtcgcg tgcacctgct ctgcgcaggg caccggcttc tccctgcccc 3000  
gcagtgtggg cgggcacccg ccccgatgc ggggtggtcac aggcgacatc ctgaccgaca 3060  
tcaccggcca caatgtctct gactacctgc tcttcacctc cgaccgcttc cgactgcacc 3120  
ggatggggc catcaccttt ggaaacgtcc tgaagtcct cccagcctca tttggacca 3180  
gggccccacc catggtgcgg aagatcgcg tgccagggc tgcccaggtt ttctacaaca 3240  
acaagggcta tcacagcatg cccacctacc tcaacgcct caacaacgcc atcctcgtg 3300  
ccaacctgcc caagagcaag ggcaacccgg cggcttacgg catcacctgc accaaccacc 3360  
ccatgaataa gaccagcgcc agcctctccc tggattacct gctgcagggg acggatgtcg 3420  
tcctgcccat ctctcatcct gtggccatgt ccttcgtgcc ggccagcttc gttgtcttcc 3480  
tcgtggcgga gaagtccacc aaggccaagc atctgcagtt tgcagcggg tgcaaccccc 3540  
tcctctactg cctggcgaa tacgtgtggg acatgtcaa ctacctggc cccgtacct 3600  
gctgtgtcat catcctgttt gtgttcgacc tgccggccta cactgcccc accaacttcc 3660  
ctgccgtcct ctccctcttc ctgctctatg ggtggtccat cagccccatc atgtaccgg 3720  
cctccttctg gttcgaggte cccagctccg cctacgtgtt cctcattgtc atcaatctct 3780  
tcctcgccat caccgcccac gtggccacct tctgtctaca gctcttcgag cactacaagg 3840  
acctgaaggt tgtcaacagt tacctgaaaa gctgcttctt cattttcccc aactacaacc 3900  
tgggccaggg gctcatggag atggcctaca acgagtacat caacgagtag tacgccaaga 3960  
ttggccagtt tgacaagatg aagtccccgt tcgagtggga cattgtcacc cgcggactgg 4020  
tgggccatggc ggttgagggc gtcgtgggct tctcctgac catcatgtgc cagtaacaat 4080  
tcctgcgggc gccacagcgc atgcctgtgt ctaccaagcc tgtggaggat gatgtggagc 4140  
tgggccagtga gcggcagcga gtgctccggg gagacgcca caatgacatg gtcaagattg 4200  
agaacctgac caaggtctac aagtccccga agattggccg tatectggcc gttgaccgcc 4260  
tgtgctggg tgtgcgtcct ggcgagtgct tcgggctcct gggcgtaaac ggtgcggggc 4320  
agaccagcac cttcaagatg ctgaccggcg acgagagcac gacggggggc gaggccttcg 4380  
tcaatggaca cagcgtgctg aaggagctgc tccaggtgca gcagagcctc ggctactgcc 4440  
cgagtggtga cgcgtgttc gacgagctca cggcccgga gcacctgcag ctgtacacgc 4500  
ggctgcgtgg gatctcctgg aaggacgagg cccgggtggg gaagtgggct ctggagaagc 4560  
tgagctgac caagtacgca gacaagccgg ctggcaccta cagcgggcgc aacaagcgga 4620  
agctctccac ggccatcgcc ctcatgggt acccagcctt catcttctg gacgagccca 4680  
ccacagggat ggaccccaag gccggcgct tctctggaa cctcactcct gacctcatca 4740  
agacagggcg ttcagtggtg ctgacatcac acagcatgga ggagtgcgag gcgtgtgca 4800  
cgcggtggc catcatgggt aacggctgcc tgcggtgcct gggcagcatc cagcacctga 4860  
agaaccgggt tggagatggc tacatgatca cgggtcgga caagagcagc cagagtgtga 4920  
aggacgtggg gcggttcttc aaccgcaact tcccgggaag catgctcaag gagcggcacc 4980  
acacaaaggt gcagtaccag ctcaagtgcg agcacatctc gctggcccag gtgttcagca 5040  
agatggagca ggtgtctggc gtgctgggca tcgaggacta ctcggtcagc cagaccacac 5100  
tggaacaatg gttcgtgaac ttgccaaga agcagagtga caacctggag cagcaggaga 5160  
cggagccgcc atccgcactg cagtcctctc tcgggtgctt gctcagcctg ctccggcccc 5220  
ggtctgcccc cacggagctc cgggcacttg tggcagacga gcccagaggc ctggacacgg 5280  
aggacgaggg cctcatcagc ttcgaggagg agcgggcccc gctgtccttc aacacggaca 5340  
cgctctgctg accacccaga gctgggcccag ggaggacacg ctccactgac caccagagc 5400  
tgggccaggg actcaacaat ggggacagaa gtccccaggt gcctgccagg gcctggagt 5460  
gaggttcagg accaaggggc ttctggtcct ccagccccgt tactcgcca tgtcctgccc 5520  
tcaactgcgtg tgccggccct 5540

&lt;210&gt; 46

&lt;211&gt; 2074



<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7472537CB1

<220>  
<221> unsure  
<222> 1869, 1898, 1988, 2000  
<223> a, t, c, g, or other

<400> 46  
ggaatcacag tgcctaggca tataataaat attcgttgaa ttaataaaat catctgatta 60  
tggtatggta gtagttcaga aaattctgtc atgaccctgt actctttctt tggaagggct 120  
ctaaatggga acaacaatat agtatgtagt ctctctgcat agctaattgt cagcaaagca 180  
gggcaatgta ggtatacaac caatctatatt ttcaactcag aaacatcaca tcatattccat 240  
tcctttataa ccatccttct tccatcccaa agtatagttt gtcaacctgg aactcaaaca 300  
ttgtatggtc tggaaatgacc gtacagtgtg aaggaggaaa agaaaattgg ggtgtcttat 360  
ttccctctct ctgattcagt tacttagatc acctgaaaca tacatatgat tcagagcata 420  
tatttagatg ttttctacttt cttatttgtg tgtgtgtgtg ttcagtcaat ttgctaata 480  
agacactgaa agtcagaaat tccctgacaaa tggatttttg gggaaaaaga agctggcaga 540  
tcccttcttt ttcaagcatc ccggaaccac ttcctttgga atgtcttcat ttaacctgag 600  
taatgccatc atgggcagtg ggatcctggg cttgtcttat gccatggcca acacagggat 660  
catacttttt atgttcatgc tgcttgcgtg ggcaatatta tcaactgtatt cagttcacct 720  
tttattaaaa acatctttga ttgtagggtc tttgatttat gaaaaattag gagaaaaggc 780  
atttgatggg ccgggaaaaa ttggagcttt tgtttccatt acaatgcaga acattggagc 840  
aatgtcaagc tacctcttta tcattaaata tgaactacct gaagtaatca gagcattcat 900  
gggacttgaa gaaacttcta gagaatggta cctcaatggc ccttaaaaaat ctaggttatc ttggctatac 1020  
gtctgttgga attattcttc cactttcgct ccttaaaaaat ctaggttatc ttggctatac 1080  
cagtggattt tctcttacct gcatgggtgt ttttggtagt gtgggtgattt acaagaaatt 1140  
ccaaataccc tgccctctac ctgagaacca ggccaagggc tctcttcatg acagtggagt 1200  
agaatatgaa gctcatagtg atgacaagtg tgaacccaaa tactttgtat tcaactccca 1260  
gacggcctat gcaattccta tccatgtatt tgcttttgta tgccaccctg aggtccttcc 1320  
catctacagt gaacttaaa atcggtcccg gagaaaaatg caaacgggtg caaatatttc 1380  
catcacgggg atgcttgta tgtacctgct tgccgcccctc tttgggtacc taaccttcta 1440  
tggtagggtt gaagatgaat tacttcatgc ctacagcaaa gtgtatacat tagacatccc 1500  
ccttctcatg gttcgcctgg cagtccttgt ggcagtaaca ctaactgtgc ccattgtcct 1560  
cttcccagtt cgtacatcag tgatcacact gttatttccc aaacgacctc tcagctggat 1620  
acgacatttc ctgattgcag ctgtgcttat tgcacttaat aatgttctgg tcactcttgt 1680  
gccaactata aaatacatct tcggattcat aggggcttct tctgccacta ggtgatttt 1740  
tattcttcca gcagtttttt atcttaaaat tggcaagaaa gaaactttta ggtcaccccc 1800  
tgaattacag gctttaattt tcttgggtgt tggaaatattc ttcatgattg gaagcatggc 1860  
actcattata attgactgga tttatgatcc tccaaattcc aagcatcact aacacaagga 1920  
aaaatactnt ctttttctat tggaaatggg tacaagtnat actccaaaag atatttgaat 1980  
tatcttgatt ggaatgttat tcataggaaa taacaggaag attccaaaaga cgtttaccag 2040  
taatatcncc aggcacctgn cagaagaggg aaaatcactg tttttgtcaa ggatggttgt 2074  
gtatgtgttt taaaataaaa cctgtggtgc acat

<210> 47  
<211> 2259  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7472546CB1

<400> 47  
atggaatacc aggcgtccga ggtgatcggg cagcgtcagt cttcagccac taagccagga 60  
agatctggga aggagtcagt cacagagccc tgggcccagag ttccaggggc tctgggagtg 120  
gctgccaggc agatgcaccc caagtcaata atcacattca gagagataaa tggggagtag 180  
actggggctg tggattttcc caggctagga gtccgtgctt ctgaggaaaac agcgtcaga 240

```

gagctgaaga tgagcaagga gctggcagca atggggcctg gagcttcagg ggacgggggtc 300
aggactgaga cagctccaca catagcactg gactccagag ttggtctgca cgcctacgac 360
atcagcgtgg tggatcatcta ctttgtcttc gtcatctgctg tggggatctg gtgctccatc 420
cgtgcaagtc gagggaccat tggcggtat ttcctggccg ggagggtccat gagctgggtg 480
ccaattggag catctctgat gtccagcaat gtgggcagtg gcttggtcat cggcctggct 540
gggacagggg ctgcccggagg ccttgccgta ggtggcttcg agtggaaacgc aacctggctg 600
ctcctggccc ttggtgggt cttcgteccct gtgtacatcg cagcaggtgt ggtcacaatg 660
ccgcagtatc tgaagaagcg atttgggggc cagaggatcc aggtgtacat gtctgtcctg 720
tctctcatcc tctacatctt caccaagatc tgcactgaca tcttctctgg agccctcttc 780
atccagatgg cattgggctg gaacctgtac ctctccacag ggatcctgct ggtgggtgact 840
gccgtctaca ccattgcagg tggcctcatg gccgtgatct acacagatgc tctgcagacg 900
gtgatcatgg tagggggagc cctggtectc atgtttctgg gctttcagga cgtgggctgg 960
taccagggcc tggagcagcg gtacaggcag gccatcccta atgtcacagt ccccaacacc 1020
acctgtcacc tcccacggcc cgatgcttcc cacattcttc gggaccctgt gaggggggac 1080
atcccttggc caggtctcat tttcgggctc acagtgtcg ccacctgggt ttggtgcaca 1140
gaccaggtca ttgtgcagcg gtctctctcg gccaaagatc tgtctcatgc caaggaggc 1200
tcctgtctgg ggggctacct gaagatcctc cccatgttct tcatcgctcat gcctggcatg 1260
atcagccggg cctgttccc agacgaggtg ggtgctgctg accctgatgt ctgcccaga 1320
atctgtgggg ccgagtgagg atgttccaac attgcctacc ctaagttggt catggccctc 1380
atgacctgtg gtctgcgggg gctgatgatt gccgtgatca tggccgctct catgagctca 1440
ctcaacctca tcttcaacag cagcagcacc ctgttcacca ttgatgtgtg gcagcgcttc 1500
cgcaggaagt caacagagca ggagctgatg gtggtgggca gagtgttgt ggtgttctg 1560
gttgtcatca gcatcctctg gatccccatc atccaaagct ccaacagtgg gcagctcttc 1620
gactacatcc aggtgtgac cagttacctg gcccaccca tcaccgctct ctctctgctg 1680
gccatcttct gcaagagggg cacagagccc ggagcttctt ggggcctcgt gtttggcctg 1740
ggagtggggc ttctgcgtat gatcctggag ttctcatacc cagcgccagc ctgtggggag 1800
gtggaccgga ggccagcagt gctgaaggac ttccactacc tgtactttgc aatcctctc 1860
tgcgggctca ctgccatcgt cattgtcatt ctacacgcc tcacatggtg gactcggaac 1920
tgccccctct ctgagctgga gaaggaggcc cacgagagca caccggagat atccgagag 1980
ccagccgggg agtgccctgc aggaggtgga gcggcagaga actcgagcct gggccaggag 2040
cagcctgaag cccaagcag gtccctggga aagtgtctct ggagctgggt ctgtgggctc 2100
tctggaacac cggagcaggg cctgagccca gcagagaagg ctgcgctaga acagaagctg 2160
acaagcattg aggaggagcc actctggaga catgtctgca acatcaatgc tgtccttttg 2220
ctggccatca acatcttct ctggggctat tttgctga 2259

```

&lt;210&gt; 48

&lt;211&gt; 2439

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7474202CB1

&lt;400&gt; 48

```

ggctctgtga gaggaggggc agttcagccg cagcaggagg actgacaggg gcctgatgga 60
ggagttgggtg gggctgcgtg agggcttctc aggggacctt gtgactctgc aggagctgtg 120
gggcccctgt ccccatatcc gccgagccat ccaaggtggc ctggagtggc taaagcagaa 180
ggtgttccgc ctgggagaag actggtactt cctgatgacc ctcggggtgc tcatggccct 240
ggtcagctat gccatgaact ttgccatcgg gtgtgtgggtc cgaggcttct cccagagcat 300
cacgcccctc tctggagggt ctggaatccc ggagctgaag accatgttgg cgggtgtgat 360
cttggaggac tacctggata tcaagaactt tggggccaag gtggtggggc tctcctgcac 420
cctggccacc ggcagcacc tgttcctggg caaagtgggc ccttctgtgc acctgtctgt 480
aatgatecgt gcctacctgg gccgtgtgcg caccacgacc atcggggagc ctgagaacaa 540
gagcaagcaa aacgaaatgc tgggtggcagc ggcggcagtg ggcgtggcca cagtctttgc 600
agctcccttc agcggcgctc tgttcagcat cgaggatcat tcttccact tctctgtccg 660
ggattactgg aggggcttct ttgcggccac ctgcggggcc ttcatatctc ggctcctggc 720
agtcttcaac agcagcagc agaccatcac ctccctctac aagaccagtt tccgggtgga 780
cgttcccttc gacctgcctg agatcttctt ttttgtggcg ctgggtggca tctgcccgt 840
cctgagctgt gcttacctct tctgtcagcg aaccttctc agcttcatca agaccaatcg 900
gtacagctcc aaactgctgg ctactagcaa gcctgtgtac tccgtctctg ccacctgtct 960
tctcgctcc atcacctacc cgcctgggtg gggccacttc ctactctctc ggctgtccat 1020
gaagcagcat ctggactcgc tgttcgacaa ccactcctgg gcgtgatga cccagaactc 1080

```

cagccccacc	tggcccgagg	agctcgaccc	ccagcacctt	tgggtgggaat	ggtaccaccc	1140
gcggttcacc	atctttggga	cccttgccct	cttcctgggt	atgaagttct	ggatgctgat	1200
tctggccacc	accatcccca	tgccctgccc	gtacttcctg	cccatcttta	tccttgaggc	1260
tgccatcggg	cgccctcttg	gagaggtctt	tgccgtcgcc	ttccctgagg	gcattgtgac	1320
tggagggggt	accaatccca	tcattgcccgt	gggtatgctt	ctggcagggg	ctgcagcctt	1380
ctcaggggct	gtgaccacaa	ccatctccac	ggcgtgctgt	gcctttgagc	tgaccggcca	1440
gatagtgcct	gcactgcccg	tgctgatggc	ggtgctggca	gccaacgcca	ttgcacagag	1500
ctgccagccc	tccttctatg	atggcaccat	cattgtcaag	aagctgccat	acctgccacg	1560
gattctgggc	cgcaacatcg	gctcccacca	tgtgaggggt	gagcacttca	tgaaccacag	1620
catcaccaca	ctggccaagg	acacgcccgt	ggaggagggt	gtcaagggtt	tgacctccac	1680
agacgtgacc	gagtatcccc	tggtggagag	cacagagtcc	cagatcctgg	taggcacgtt	1740
gcagagggct	cagctgggtg	aggccctcca	ggctgagcct	ccttcagggg	ctccaggaca	1800
ccagtgtctc	caggacatct	tggccagggg	ctgccccacg	gaaccagtga	ccctgacgct	1860
attctcagag	accaccttgc	accaggcaca	aaacctcttt	aagctgttga	accttcagtc	1920
cctcttcgtg	acatcgcggg	gcagagctgt	gggctgcgtg	tcctgggtgg	agatgaagaa	1980
agcaatttcc	aacctgacaa	atccgccagc	tccaaagtga	gccggcccag	caagatgaaa	2040
cagggcaccc	cagctgacct	ggtactgagg	ttgggctgag	acctgcttcg	tcctcccca	2100
tcaccacctg	cccctccctc	cagcccagct	ccattctttg	gcataacagg	caactctaag	2160
ctagcccaga	agaggatggc	tcactctggg	tgggacgatg	gctcctgctt	tgaagacaaa	2220
aaatcccacc	ttgggagag	ctgagtgtga	gaagatggaa	aaccagtatc	tgccagggtg	2280
tcagtgactg	gccatcacat	taatgaatga	cgagattgga	gtacactgtc	accaagggca	2340
ggcaaagatg	ccctctgggg	ttgtctgggt	cccagtgaga	ggctcctgag	aaaaataaag	2400
ctggttccca	gagctgctgt	ccatccctca	aaaaaaaa			2439

&lt;210&gt; 49

&lt;211&gt; 2762

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7476280CB1

&lt;400&gt; 49

atggacccca	tcacgcctaa	ctggactgag	atcgtgaaca	ggaagctcag	cttcccacct	60
ccactcctgg	atgccatcca	ggagggccga	ctgggctttg	tgacgagct	gctggagtca	120
gaggttgagg	ccgcagcag	tgggcccagg	tggccctgtt	ggaatgtgga	agaggctgag	180
gaccgctgct	ggagggaggc	actcaacctg	gccatccgcc	tgggcatga	ggccctcacc	240
gatgtgctgt	tggccagtgt	caagtttgac	ttccgccaga	tccatgaggc	cctgctagtg	300
gcagtggaca	caaaccaggc	agtgggtgct	cgccctgccc	ccgggctgga	acgggagaag	360
ggtcgcaaag	tagacaccag	gtctttctca	ctggctttct	ttgactcctc	aattgatggc	420
tccegccttg	cacctgggtg	gactccccct	ccccaggcct	gccagaagga	cctgtatgag	480
atagcacagg	tgctcatgga	acagggccac	accattgccc	ggccccacce	ggtctcctgt	540
gcctgcctcg	agtgcagcaa	cgcccgccgc	tatgacctgc	tgaactcttc	tctgtcccgc	600
atcaacacct	accttggcat	cgccagcagg	gccacctctt	cactggccag	tgaggatgcc	660
atgctggctg	ccttccagct	tagccgtgag	ctcaggcgcc	ttgcaacgaa	ggagcctgaa	720
tttaagcctg	agtacattgc	tctggagtca	ctgagccagg	actatggctt	tcagctgctg	780
ggcatgtgct	ggaaccagag	tgaggtcact	gcagtgtcca	acgacctggc	cgaggacagc	840
gagactgagc	ccgaggtgga	aggcctgggc	ctggcctttg	aggaaggcat	ccccaacctg	900
gtgaggctgc	gactggctgt	caactacaac	cagaagcggg	tcgtagcaca	cctcatctgc	960
cagcaagtcc	tgtctctccat	ctgggtgtgg	aacctggctg	gttggcgggg	aagcaccacc	1020
agctggaagc	tccttgctac	cttctctcat	ttcctcacca	tgcccttctt	ctgccttggc	1080
tactggctga	caccaaagtc	ccagctgggc	cacctgctaa	agatcccagt	actgaagttc	1140
ctgctgcact	ctgcctccta	tctgtgggtt	ctcatcttcc	tgctgggaga	gtccctggtc	1200
atggagacac	agctgagcac	cttccgtggc	cgcagccaga	gtgtctggga	gacttcacta	1260
cacatgattt	gtgtcacagg	cttctctgtg	tttgagtgca	aggaagtgtg	gattgagggc	1320
ctgcgcagtt	acctcctgga	ctgggtggaac	ttcctgggata	tggtcgtcct	gtccctgtac	1380
ctggcagcct	tcgcaactgc	cctctctctg	gctgggcttg	ccccatgca	ctgccgggac	1440
gcctcccaag	cggctgcctg	ccactatttc	accatggctg	aaagaagcga	gtggcacacc	1500
gaggatcccc	agttcttggc	tgaggtgctc	ttcactgcca	ccagcatgct	cagcttcacc	1560
cgccctggct	acattctgcc	ggcccaacag	tcgctgggca	ctctgcagat	ttccattggc	1620
aagatgattg	aagacatgat	cgggtttatg	ttcatcctca	tgatcatcct	gaccgccttc	1680
ctctgtggcc	tcaacaacat	ctatgtgccc	taccagaaga	cagagtggct	gggcaagagt	1740

```

ttcaatgaga cgttttcagtt tctgtttctgg accatgtttcg gtatggaaga gcacagcgtg 1800
gtggacgtgc ctcagttttct ggtgccccgag tttgcaggcc gggccctcta tggcatcttt 1860
accatcatca tggtcattgt gctgctcaac atgctcattg ctatgatcac caactccttc 1920
cagaagattg aggatgatgc tgacgtggag tggacgtttg ctgctccaa gctgtatctg 1980
ttctacttcc gagagggcct gacactgcct gtgcccttca acatcctgcc ctectogaag 2040
gctgtcttct accttctcag gagaatttgc cagttcattt gctgttgctg ttcctgctgc 2100
aaaaccaaga agccagacta tcccccgatc cctacttttg tgaatcccag ggcaggggct 2160
gtgcctgggg agggagagcg tggatcctac cgccttcacg tcatcaaggc cctggtacag 2220
cgctacacag agactgccc gcgagaattc gaggagacc ggcgaaaga tctgggcaac 2280
agactcacag agctgaccaa gaccatatct cgactgcaaa gcgaggtagc cgtgtgctgc 2340
agaactctgg cagagggagg gacgccccgg cctcccgacg gtgccagcgt cctcagtcac 2400
tacatcactc aagtgcacaa cagcttccag aacctggggc ctcccatccc tgagacccca 2460
gagctgacag ggcctgggat tgtgaggacc caggaatcat caggaaccgg gcttcaggac 2520
actggagggg tgaggactct ggcttcggga gagtctggcc cctgctcccc agctcatgtg 2580
ctagttcata gggagcagga agcagagggg gctggggacc tgccccaggg ggaggattcg 2640
gggactgaga ggaggtcctg atacagtga agagtccctt cttctgttgc tgagcgtggg 2700
agcctaggag ggtgaggggtg gggggccctt tgggaggagc ctgtgctgct tttcttgctt 2760
ca
2762

```

&lt;210&gt; 50

&lt;211&gt; 1897

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1713377CB1

&lt;400&gt; 50

```

gcatctaga actagtgagc tgcaggctgg catggctggg gggatgtcag cggagtggcc 60
tgagcctggg ccaggaggctc tgcagggcca gtccccaggg ccaggcaggc agtgtcccc 120
tcccatcacg cccacctcct ggagcctgcc cccgtggagg gcctacgtgg ctgccgccgt 180
cctctgctac atcaacctcc tgaattacat gaactgggtc atcattgcag gagtgtgct 240
ggatatacac gaggttttcc agatcagtga caaccatgct ggtttgcttc agatgtctt 300
cgtagctgc ctgctgctgt ctgcacctgt gtttggttac ctgggcgacc gacatagccg 360
caaggctacc atgagcttcg gtatcttgct gtggtcagga gctggcctct ctagctcctt 420
catctcccc cggtattctt ggctcttctt cctgtcccgg ggcacgtgg gcactggctc 480
ggccagctac tccaccatcg cgcccaccgt cctgggcgac ctcttcgtga gggaccagcg 540
caccgcgtg ctggctgtct tctacatctt tatccccgtt ggaagtggte tgggctacgt 600
gctggggctg gctgtgacga tgctgactgg gaactggcgc tgggcccctc gagtcatgcc 660
ctgcctggag gccgtggcct tgatcctgct tatectgctg gttccagacc caccocgggg 720
agctgccgag acacaggggg agggggccgt gggaggcttc agaagcagct ggtgtgagga 780
cgtcagatac ctggggaaaa actggagttt tgtgtggctg accctcggag tgaccgccat 840
ggcctttgtg actggagccc tggggttctg ggcccccaag tttctgctcg aggcacgcgt 900
ggttcacggg ctgcagcctc cctgcttcca ggagccgtgc agcaaccccg acagcctgat 960
ttttggggca ctgacatca tgaccggcgt cattgggggtc atcttggggg cagaagcttc 1020
gaggaggtag aagaaagtca ttccaggagc tgagcccctc atctgcccct ccagcctgct 1080
tgccacagcc cctgcctct acctggctct cgtcctggcc ccgaccacc tgctggcctc 1140
ctatgtgttc ctgggccttg gggagctgct tctgtcctgc aactgggcag tggttgccga 1200
catcctgctg tctgtggtgg tgcccagatg ccgggggacg gcagaggcac ttcagatcac 1260
ggtggggcac atcctgggag acgtggcag cccctatctc acaggactta tctctagtgt 1320
cctgcggggc aggcgcctg actcctatct gcagcgttc cgcagcctgc agcagactt 1380
cctgtgctgc gectttgtca tgcctctggg ggcggctgc ttcctgctga ctgcgtgta 1440
cctggagaga gacgagacc gggcctggca gcctgtcaca gggacccag acagcaatga 1500
tgtggacagc aacgacctgg agagacaagg cctactttcg ggcgctggcg cctctacaga 1560
ggagccctga ggtccctgcc tacactcgtc ctgctgcaa gcctcccgtt ggtccccaca 1620
gcagcagtg ctcggttcct ctttggctgt cctcggggac tccggttag gcacatctgc 1680
cacttttgaa ttcccggctg gagagctggc aggacctgt ggctgggctg ggaatggagc 1740
tgtcagcact tccgctggga ggctggggc tgtgctgca tcccgtcaa ggctgcccc 1800
gcctggggtc tccagcctgg ctgctgctgg gccctgaata aagagaggcc agtacaaagc 1860
ccatggattt tgggcctgta aaaaaaaaa aaaaaaa 1897

```

&lt;210&gt; 51

<211> 2361  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 5842557CB1

<400> 51  
gatgatggca gacaggagag ctgactactt tcagaacctg cctgagtctc tgacttccct 60  
tcttggtgct tctgaccacg tccaacaacc ccgatgtgat gattcctgcg tattccaaga 120  
accgggccta tgccatcttc ttcatagtct tcaactgtgat aggaagcctg tttctgatga 180  
acctgctgac agccatcatc tacagtcagt tccggggcta cctgatgaaa tctctccaga 240  
cctcgctgtt tcggaggcgg ctgggaaccc gggctgcctt tgaagtccta tcctccatgg 300  
tgggggaggg aggaagcctt cctcaggcag ttgggggtgaa gccccagaac ttgctgcagg 360  
tgcttcagaa ggtccagctg gacagctccc acaaacaggc catgatggag aaggtgcgtt 420  
cctacgacag tgttctgctg tcagctgagg agtttcagaa gctcttcaac gagcttgaca 480  
gaagtgtggt taaagagcac ccgcccaggc ccgagtacca gtctccgttt ctgcagagcg 540  
cccagttcct ctccggccac tactactttg actacctggg gaacctcatc gccctggcaa 600  
acctggtgtc catttgctg ttcctgggtgc tggatgcaga tgtgctgcct gctgagcgtg 660  
atgacttcat cctggggatt ctcaactgcg tcttcattgt gtactacctg ttggagatgc 720  
tgctcaagggt ctttgccctg ggctgagag ggtacctgtc ctaccccagc aacgtgtttg 780  
acgggctcct caccgttgte ctgctggttt tgggctgctc aactctgggt gtgtaccgat 840  
tgccacacct aggtgagg cggagatgg tgggctgct gtcgctgtgg gacatgacct 900  
gcatgctgaa catgctcatc gtgttccgt tctgctgat catcccagc atgaagccga 960  
tggccgtggt ggccagtacc gtctgggccc tgggtgcagaa catgctgct tttggcggga 1020  
tcttggtggt ggtctactac gtatttgcca tcattgggat caacttgttt agaggcgtca 1080  
ttgtggctct tcttggaac agcagcctgg cccctgccaa tggctcggcg cctgtggga 1140  
gcttcgagca gctggagtac tgggccaaca acttcgatga ctttgccgggt gccctgggtca 1200  
ctctgtggaa cttgatggtg gtgaacaact ggcaggtgtt tctggatgca tatcggcgtc 1260  
actcaggccc gtggtccaag atctattttg tattgtggtg gctggtgtcg tctgtcatct 1320  
gggtcaacct gtttctggcc ctgattctgg agaacttctc tcacaagtgg gacccccgca 1380  
gccacctgca gcccttctgct gggacccag aggcaccta ccagatgact gtggagctcc 1440  
tgctcaggga tattctggag gagcccaggg aggatgagct cacagagagg ctgagccagc 1500  
acccgcacct gtggtgtgac aggtgacgct cgggctgccc tcccagcagg ggcggcagga 1560  
gagagagggt ggcctacaca ggtgcccgct atggaagagg cggccatgct gtggccagcc 1620  
aggcaggaa agacctttcc tctgacggac cactaagctg gggacaggaa ccaagtcctt 1680  
tgctgtggc ccaacaaccg tctacagaac agctgctggt gcttcaggga ggcgcgtgc 1740  
cctccgcttt cttttatagc tgcttcagtg agaattccct cgtcgactcc acaggggacct 1800  
ttcagacaaa aatgcaagaa gcagcggcct cccctgtccc ctgcagcttc cgtggtgcct 1860  
ttgctgccgg cagcccttgg ggaccacagg cctgaccagg gctgcacag gtttaaccgtc 1920  
agacttccgg ggcattcagg tggggatgct ggtggtttga catggagaga accttgactg 1980  
tgcttctggt tttcatcaat aagtcttgca aagaatggga ttgtcattct tcaactcagc 2040  
acagttctag tctgcttct ctggagtagg gttgttgagt aaggttgctt ggggttggtgca 2100  
tttgacaaag ggcacatggc tgtgaggtgt atcctggcgg ggggctgtct acctgcagtgc 2160  
aggggcacct tttctgtttt gctcaaaggc atgtataagc caatgggtga ccttatttcc 2280  
tgtgtcttca ggtgtgtgca ggggcctggg gtggggagtt gggggagcga gcagtgtgtg 2340  
gaaggggatc cactagtctt a 2361

<210> 52  
<211> 2032  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7476643CB1

<400> 52  
gccttggcag agtctggggt ccctggactg agccatcagc tgggtcactg agaccatagg 60  
caaggaaaca aaataggaat tccaaggaac tgggcctagt tcccctcaca gatgacacca 120  
gccacgccag gcctccaggg ccaggagggg cactgctgga gtgtgaccac ctgaggagtgc 180

```

gggtgccagg tgggaaggaga agaaaggact ggtcctgctc gctcctcgtg gcctccctcg 240
cggggcgctt cggctcctcc ttcctctacg gctacaacct gtcgggtggtg aatgccccca 300
ccccgtacat caaggccttt tacaatgagt catgggaaag aaggcatgga cgtccaatag 360
accagacac tctgactttg ctctggtctg tgactgtgtc catattcgcc atcgggtggac 420
ttgtggggac gttaattgtg aagatgattg gaaagggttct tgggaggaag cacactttgc 480
tggccaataa tgggtttgca atttctgctg cattgctgat ggcctgctcg ctccaggcag 540
gagcctttga aatgctcatc gtgggacgct tcatcatggg catagatgga ggcgtcgccc 600
tcagtgtgct ccccatgtac ctcatgaga tctacccaa ggagatccgt ggctctctgg 660
ggcaggtgac tgccatcttt atctgcattg gcggtgtcac tgggcagctt ctgggacctg 720
ccgagctgct gggaaaggag agtacctggc catacctgtt tggagtgatt gtggctccctg 780
ccgtttgcca gctgctgagc cttccctttc tcccggacag cccacgctac ctgctcttgg 840
agaagcaca cagaggcaaga gctgtgaaag ccttccaaac gttcttgggt aaagcagacg 900
tttcccaaga ggtagaggag gtcttggctg agagctccct acgtccgctg gcagaggagc atccgcctgg 960
tgtccgtgct ggagctgctg agagctccct acgtccgctg gcaggtgggt accgtgattg 1020
tcacatggc ctgctaccag ctctgtggcc tcaatgcaat ttggttctat accaacagca 1080
tctttggaaa agctgggac cctctggcaa agatccata cgtcaccttg agtacagggg 1140
gcatcgagac tttggctgcc gtcttctctg gtttgggtcat tgagcacctg ggacggagac 1200
ccctcctcat tgggtggcttt gggctcatgg gcctcttctt tgggaccctc accatcacgc 1260
tgaccctgca ggaccacgcc cctgggtcc cctacctgag tatcgtgggc attctggcca 1320
tcacgcctc tttctgcagt gggccagggt gcatcccggt catcttgact ggtgagttct 1380
tccagcaatc tcagcggccg gctgccttca tcattgcagg caccgtcaac tggctctcca 1440
actttgctgt tgggctcctc ttccatttca ttcagaaaag tctggacacc tactgtttcc 1500
tagtctttgc tacaatttgt atcacagggt ctatctacct gtattttgtg ctgcctgaga 1560
ccaaaaacag aacctatgca gaaatcagcc aggcattttc caaaaggaa aaagcatacc 1620
caccagaaga gaaaatcgac tcagctgtca ctgatgtca aaggaactaa gacaaagatc 1680
atggagacca tcgggtgagt ctcaagact ccccgagctc tgcttggctg gtctcctgct 1740
ggtaattttc gtctgtagag aggaacaaga acttccatt tatcttgctt acctgcactt 1800
atgaaaagtc aaactgagtc atgctgagag ccagggaaca taggagtcag ttcttctgca 1860
gcagcactca gccagttgaa ggcaatgtgg agtgatggaa ggagagcagt gatgcagtga 1920
tgctggcacc aactccttta ctatggcatc cattgtacca gctgccatac accaggcaac 1980
ttctacactt tatctcta ctcctagaa taagtattag tttcccccac tt 2032

```

&lt;210&gt; 53

&lt;211&gt; 2779

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7611651CB1

&lt;400&gt; 53

```

cgctgtggc tccgggcagg ggcgcgggcc gaaagatgcc ggtccgcagg ggccacgtcg 60
ctccccaaa cacttacctg gacaccatca tccgcaagtt cgagggccaa agtcggaagt 120
ctctgattgc caatgctcag atggagaact gcgccatcat ttactgcaac gacggcttct 180
gcgaactctt cggctactcc cgagtggagg tgatgcagca accctgcacc tgcgaacttc 240
tcacaggccc caacacacca agcagcgccg tgtcccgct agcgcaggcc ctgctggggg 300
ctgaggagtg caaggtggac atcctctact accgcaagga tgccctcagc ttccgctgcc 360
tggtagatgt ggtgcccgtg aagaacgagg acggggctgt catcatgttc attctcaact 420
tcgaggacct ggcccagctc ctggccaagt gcagcagccg cagcttgtcc cagcgccgtg 480
tgtcccagag cttcctgggc tccgagggt ctcatggcag gccaggcgga ccagggccag 540
gcacaggcag gggcaagtac aggaccatca gccagatccc acagttcacg ctcaacttcg 600
tggagttcaa cttggagaag caccgtcca gctccaccac ggagattgag atcatcgcg 660
cccataaggt ggtggagcgg acacagaacg tcatgagaa ggtcaccag gtctgtccc 720
tgggcgcgga tgtgctgccg gactacaagc tgcaggcgcc ggcacccac cgctggacca 780
tctgcaacta cagccccctt aaggccgtgt gggactggct catcctgctg ctggtcatct 840
acacggctgt cttcacgccc tactcagccg ccttctgct cagcgatcag gacgaatcac 900
ggcggtgggc ctgcagctat acctgcagtc cctcactgt ggtggatctc atcgtggaca 960
tcatgttcgt cgtggacatc gtcatacaact tccgcaccac ctatgtcaac accaatgatg 1020
aggtggctcag ccacccccgc cgcacgcgg tccactactt caagggtgg ttccctcattg 1080
acatgggtggc cgccatccct ttgcacctcc tgatcttccg cactggctcc gatgagacca 1140
caaccctgat tgggtatttg aagacagcgc ggctgctgcg gctgggtgcg gtagcacgga 1200
agctggaccg ctactctgag tatggggcgg ctgtgctctt cttgctcatg tgcaccttcg 1260

```

cgctcatagc	gcactggctg	gcctgcatct	gcagcctcac	cagcgtgggc	ttcggcaatg	1320
tctcgcccaa	caccaactcc	gagaagggtc	tctccatctg	cgctcatgctc	atcgggtccc	1380
tgatgtacgc	cagcatcttc	gggaacgtgt	ccgcgatcat	ccagcgccctg	tactcgggca	1440
ccgcgcgcta	ccacacgcag	atgctgctgt	tcaaggagtt	catccgcttc	caccagatcc	1500
ccaacccact	gcgcagcgc	ctggaggagt	atttccagca	cgcttggtcc	tacaccaatg	1560
gcattgacat	gaacgcggtg	ctgaagggtc	tccccagtg	cctgcaggct	gacatctgcc	1620
tgcacctgca	ccgcgcactg	ctgcagcact	gcccagcttt	cagcggcgcc	ggcaagggtc	1680
gcctgcgctc	gctagccgtc	aagttcaaga	ccacccacgc	gccgcctggg	gacacgctgg	1740
tgcacctcgg	cgacgtgctc	tccacctctc	acttcatctc	ccgaggctcc	atcgagatcc	1800
tgcgcgacga	cggtggtcgtg	gccatccctag	gaaagaatga	catctttggg	gaacccgtca	1860
gcctccatgc	ccagccaggc	aagtcagtg	cagacgtgcg	ggctctgacc	tactgcgacc	1920
tgcacaagat	ccagcgggca	gatctgctgg	agggtgctgga	catgtaccgc	gcctttgcgg	1980
agagcttctg	gagtaagctg	gaggtcacct	tcaacctgcg	ggacgcagcc	gggggtctcc	2040
actcatcccc	ccgacaggct	cctggcagcc	aagaccacca	aggtttcttt	ctcagtgaca	2100
accagtcaga	tgcagccctc	cccctgagca	tctcagatgc	atctggcctc	tggcctgagc	2160
tactgcagga	aatgccccca	aggcacagcc	cccaaagccc	tcaggaagac	ccagattgct	2220
ggcctctgaa	gctgggctcc	aggctagagc	agctccaggc	ccagatgaac	aggctggagt	2280
cccgcgtgtc	ctcagacctc	agccgcactc	tgcagctcct	ccagaagccc	atgccccagg	2340
gccacgccag	ctacattctg	gaagcccctg	cctccaatga	cctggccttg	gttcctatag	2400
cctcggagac	gacgagtcca	gggcccaggc	tgccccaggg	ctttctgcct	cctgcacaga	2460
ccccaaagta	tggagacttg	gatgactgta	gtccaaagca	caggaactcc	tccccaggga	2520
tgcctcacct	ggctgtggca	atggacaaaa	ctctggcacc	atcctcagaa	caggaacagc	2580
ctgaggggct	ctggccaccc	ctagcctcac	ctctacatcc	cctggaagta	caaggactca	2640
tctgtggtcc	ctgcttctcc	tccctccctg	aacaccttgg	ctctgttccc	aagcagctgg	2700
acttccagag	acatggctca	gatcctggat	ttgcagggag	ttggggccac	tgaactccaa	2760
gataaagaca	ccatgagggg					2779

&lt;210&gt; 54

&lt;211&gt; 2430

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2522075CB1

&lt;400&gt; 54

atggccgagg	ccgcggagcc	ggaggggggtt	gccccgggtc	cccagggggc	gccggagggtc	60
cccgcgcctc	tggctgagag	acccggagag	ccaggagccg	cgggcgggga	ggcagaagggt	120
ccggagggga	gcgagggcgc	agaggaggcg	ccgaggggcg	ccgccgctgt	gaaggaggca	180
ggagggcgcg	ggccagacag	gggcccggag	gccgaggcgc	ggggcacgag	gggggcgcac	240
ggcgagactg	aggccgagga	gggagccccc	gagggtgccg	aggtgcccc	aggaggggag	300
gagacaagcg	gcgcgcagca	ggtggagggg	gcgagcccg	gacgcggcgc	gcagggcgag	360
ccccgcgggg	aggctcagag	ggagcccag	gactctgcgg	cccccgagag	gcaggaggag	420
gcggagcaga	ggcctgaggt	cccgaaggt	agcgcgtccg	gggagggcgg	ggacagcgta	480
gacgcggagg	gcccgcctgg	ggacaacata	gaagcggagg	gcccggcggg	cgacagcgta	540
gaggcggagg	gcccgggtgg	ggacagcgta	gacgcggaag	gtccggcggg	ggacagcgta	600
gacgcggagg	gcccgcctgg	ggacaacata	caagccagag	gcccggcggg	ggacagcgta	660
gacgcggagg	gcccgggtgg	ggacagcgta	gagggcgagg	gtccggcggg	ggacagcgta	720
gacgcggagg	gcccgggtgg	ggacagcgta	gagggcgagg	gtccggcggg	ggacagcgta	780
gacgcggagg	gcccgggtgg	ggacagcgta	gagggcgagg	gtccggcggg	ggacagcgta	840
gacgcggagg	gcccgggtgg	ggacagcgta	gagggcgagg	gtccggcggg	ggacagcgta	900
gacgcggagg	gcccgggtgg	ggacagcgta	gagggcgagg	gtccggcggg	ggacagcgta	960
gacgcggagg	gcccgggtgg	ggacagcgta	gagggcgagg	gtccggcggg	ggacagcgta	1020
gacgcggagg	gcccgggtgg	ggacagcgta	gagggcgagg	gtccggcggg	ggacagcgta	1080
gacgcggagg	gcccgggtgg	ggacagcgta	gagggcgagg	gtccggcggg	ggacagcgta	1140
gacgcggagg	gcccgggtgg	ggacagcgta	gagggcgagg	gtccggcggg	ggacagcgta	1200
gacgcggagg	gcccgggtgg	ggacagcgta	gagggcgagg	gtccggcggg	ggacagcgta	1260
gacgcggagg	gcccgggtgg	ggacagcgta	gagggcgagg	gtccggcggg	ggacagcgta	1320
gacgcggagg	gcccgggtgg	ggacagcgta	gagggcgagg	gtccggcggg	ggacagcgta	1380
gacgcggagg	gcccgggtgg	ggacagcgta	gagggcgagg	gtccggcggg	ggacagcgta	1440
gacgcggagg	gcccgggtgg	ggacagcgta	gagggcgagg	gtccggcggg	ggacagcgta	1500
gacgcggagg	gcccgggtgg	ggacagcgta	gagggcgagg	gtccggcggg	ggacagcgta	1560

gatgtgaata	agatcgagga	gttcttagag	gagaaattag	ctcccccgag	gtatcccaag	1620
ctggggaccc	aacatcccga	atctaattcc	gcaggaaatg	acgtgtttgc	caaattctca	1680
gcgtttataa	aaaacacgaa	gaaggatgca	aatgagattc	atgaaaagaa	cctgctgaag	1740
gccctgagga	agctggataa	ttacttaa	agccctctgc	ctgatgaaat	agatgcctac	1800
agcaccgagg	atgtcactgt	ttctggaagg	aagtttctgg	gtggggacga	gctgacgctg	1860
gctgactgca	acctcttacc	caagctccat	attattaaga	ttgtggccaa	gaagtacaga	1920
gattttgaat	ttccttctga	aatgactggc	atctggagat	acttgaataa	tgcttatgct	1980
agagatgagt	tcacaaatac	gtgtccagct	gatcaagaga	ttgaacacgc	atattcagat	2040
gttgcaaaaa	gaatgaaatg	aagctgggct	gttttctgtc	ttatttctca	gttgagtga	2100
caaggatacg	aaaacagtgt	gtttgaaaac	aaattaggtt	tgggttcaat	tccttcaatt	2160
tttaaaaaac	tggtctctga	gagtttttta	aatcattgag	agcctgtttt	tcttctctaa	2220
aacattagtt	taattttctt	caaaatgaaa	atactgcttt	gtaattacaa	aatgagacac	2280
acctatcttg	atatttttaa	gcaatatcag	agggtgtaaa	gaaggacatt	ttaacaatcg	2340
ccttcaattt	tactccactt	aattaccgaa	aacttactgg	agaacatggt	ccaaatcttc	2400
agtatcttgt	tctctctctc	tctctctctc				2430